A FLOW MODEL BASED, CONTROL-THEORETIC APPROACH TO DYNAMIC ROUTING IN SOME STRUCTURED COMMUNICATION NETWORKS

by

M. V. VINODKUMAR

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

May, 1996

A Flow Model Based, Control-Theoretic Approach to Dynamic Routing in Some Structured Communication Networks

A Thesis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

M.V.VINODKUMAR

to the

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

MAY, 1996

CERTIFICATE

It is certified that the work contained in the thesis entitled *A Flow Model Based, Control-Theoretic Approach to Dynamic Routing in Some Structured Communication Networks* by **M. V. Vinodkumar**, has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

Dr. S.K.Bose
Professor
Department of Electrical Engineering,
Indian Institute of Technology, Kanpur.

Dr. P.R.K. Rao
Professor
Department of Electrical Engineerin
Indian Institute of Technology, Kanpu

MAY 1996

SYNOPSIS

In this thesis, the problem of synthesising optimal dynamic routing strategies fc some specific network topologies is investigated. This investigation has bee carried out within the framework of dynamic flow models.

Traffic routing and related areas of network management are currently under going review in many of the existing networks. The concept of dynamic routing which allows for the continuous updation of the routing tables depending upon the instantaneous traffic conditions and the network status, is aimed at profiting from the existence of spare capacities in parts of the network while other parts are over loaded. Dynamic routing strategies are known to offer significant improvements in the efficiency over static routing strategies, provided a proper exploitation of the traffic variations is carried out. The main limitation with dynamic routing has been that it requires a large amount of data to be processed to perform the complex routing calculations. This, however, has been substantially overcome in the recent times, due to the continued improvements in computer technology. Therefore there is a shift in trend towards implementing dynamic routing strategies in both packet-switched and circuit-switched networks. State dependent routing schemes for telephone traffic, in which the route for a call is chosen on the basis of the occupancy-state of trunk groups at call arrival, have received much attention in the last few years [1].

Along with this shift in trend towards dynamic routing, there has also been a growing need to have theoretical frameworks within which, the problem of routing, more specifically that of dynamic routing, can be formulated. While traditionally traffic phenomena in communication networks have been analysed with queueing models [2], these traditional analytical techniques are usually inadequate in dealing with networks with state sensitive dynamics. Queueing analysis is often found to be intractable under non-stationary traffic conditions. In the context of routing, these descriptive queueing models tell us something at best about the performance of a network with a specified routing strategy. The task of synthesising the best

routing strategy (or at least a good strategy) for a given system is rarely addressed to.

Alternative traffic models, some of which at least are prescriptive [4] in nature, have been suggested in the context of traffic control, in the recent times. A significant contribution in terms of providing a modelling scheme which allows the synthesis of advanced traffic control rules within the framework of optimal control theory was made by J. Filipiak [3]. The approach adopted in [3] is to model the network by dynamic flow models. The basic model of dynamic flows relates the growth in the amount of packages/messages in the system, by means of a deterministic differential equation, to the intensity of newly arriving traffic, intensity of traffic discarded by the system and the intensity of successfully delivered traffic. It was assumed by Filipiak in [3] that the intensity of outgoing traffic can be approximated by a (non-linear) function of the system state. The work in this thesis starts with this framework.

In a packet switching context, it is reasonable to assume that the flow out from any buffer onto its associated link will increase with increasing buffer occupancy, and will saturate at a value equal to the channel capacity of the link. It is also reasonable to assume that the flow out will be zero when the buffer is empty. Based on these considerations, we concentrate in this thesis, on a model in which the flow out function increases linearly with the buffer occupancy and saturates at a value equal to the channel capacity of the link. Apart from these practical reasons, as against more complex form of functions for the flow out given in [3], we consider that, this class of models (in which the flowout increases linearly with buffer occupancy and saturates at the channel capacity of the link) will make an interesting study in itself. The emphasis in our work is therefore on this model.

The possibility of treating the network as a dynamical system with the routing variables as control variables, allows us to formulate the problem of optimal dynamic routing as an optimal control problem. This problem is solved by a well known technique of optimal control theory, namely Pontryagin's Maximum Principle. The necessary conditions the optimal solution has to satisfy are specified in

terms of a two point boundary value problem in the state and costate variables. It is well known that it is not easy to obtain the solution to these equations. The problem becomes more complicated when the dimension of the network is large, since the dimensionality of the system of differential equations in the state and costate variables also increases correspondingly. The emphasis in Filipiak's work [3] is to consider the steady-state solution to the costate variables when the duration of network operation tends to infinity. Under this assumption, the solution to the costate variables (steady state) can be obtained by solving a set of algebraic equations.

In this thesis, we relax the assumption of an infinite time duration operation of the network. Since in practice, networks are subjected to shut downs and need rebooting from time to time, we consider that the assumption of a finite time duration of operation corresponds to a practically more realistic situation.

An optimal dynamic routing strategy is defined as one which minimises the total buffer occupancy time. The rationale for choosing this is that a waiting cost is incurred at a rate proportional to the number of customers in the system.

Since large networks can be viewed as being composed of simpler network structures, the problem of synthesising optimal (or at least good suboptimal) routing strategies for these large networks can be approached by synthesising them for the simpler network units. With this perspective, we investigate the problem of optimal routing in two simple structures namely a two node unit and a three node unit in this thesis. We also consider some network topologies which are composed of these units. The main results regarding the nature of optimal routing strategy for these network structures are summarised below.

Main Results

The thesis is organised into six chapters. In Chapter 1, we give a control-theoretic perspective to routing and state the motivations as well as the contributions of this thesis.

In Chapter 2 we review the various modelling schemes used in the performance study of networks. We review the dynamic flow model proposed by Filipiak [3] in detail and highlight the reasons for the choice of this modelling scheme in the synthesis of traffic control rules. A brief survey of routing strategies employed both in the context of circuit switched networks and packet switched networks is also presented in this chapter.

In Chapter 3, we investigate the problem of optimal routing in a two node network, in which the *faster* link has a finite channel capacity. Under the assumption that the initial buffer occupancy on the *faster* link is below the saturation value (we argue that this assumption, though may appear to be overly restrictive, is not quite so in practice), we derive the set of equations in terms of the link parameters and the input traffic, the solution to which specifies the optimal routing strategy. Solution to these equations requires the knowledge of the load pattern for the entire duration of network operation, and this in turn necessitates an off-line computation. We therefore propose an on-line implementable suboptimal routing strategy for this network. Some numerical examples comparing the performances of the optimal and suboptimal strategies are given.

In Chapter 4, we investigate the problem of optimal routing in a three node network. We start with the assumption that all the links of the network have infinite channel capacities. The optimal routing strategy in this case is *bang-bang* and can be completely specified in terms of a single switching instant. The equation that specifies this switching instant is also obtained in terms of the link parameters of the network and the duration (T) of operation of the network. It is shown that the optimal routing strategy has the *loop-free* property.

We then relax the assumption of infinite channel capacities and consider the case wherein one of the direct links has finite channel capacity. We show that the *loop-free* property holds good under this situation also. The routing strategy is not *bang-bang* in nature (we provide examples wherein the routing variables take non-zero, non-unity values over intervals). The various modes of network operation are defined and it is shown that out of the 27 possible modes of operation of the network, the

optimal routing strategy admits only four of these in a terminal interval. These cases are illustrated with examples.

We then impose certain additional assumptions on the link parameters of the network, which in practical terms imply that corresponding to a situation in which all the links have infinite channel capacities, the optimal routing strategy is to route the entire traffic arriving at a source node onto the direct link for the entire duration of network operation. Under the assumption that the initial buffer occupancy on the link with finite channel capacity is below the saturation value, the set of equations (in terms of the link parameters and the input traffic) required to specifies the optimal routing strategy is derived. Solution to these equations requires the knowledge of the load pattern for the entire duration of network operation and this in turn necessitates an off-line computation of the routing strategy. As in the case of Chapter 3, we propose an on-line implementable suboptimal strategy for this network. Some numerical examples comparing the performance of the optimal and suboptimal routing strategies are given.

In the final section of Chapter 4, we investigate the case wherein the flow out on a link is an exponential function of the (associated) buffer occupancy. The optimal routing strategy in this case is obtained by numerically solving the two point boundary value problem in the state and costate variables. From the numerical investigations carried out for various choice of link parameters, initial buffer occupancies and input traffic, we conjecture that the optimal routing strategy has the *loop-free* property. It is also observed that there is at most one switching instant and that the network operation always ends with a direct routing of packets at both the source nodes.

In Chapter 5, we investigate the problem of optimal routing in some specific networks which can topologically be viewed as being composed of the two node network units and the three node network units. We first consider a network which is a cascade of n two node units considered in Chapter 3. For the case wherein m of such network units have their *faster* link of finite channel capacity (both the links of the remaining $(n-m)$ units are assumed to be of infinite channel capacity)

we study the properties of the optimal routing strategy. It is proved that in all the network units which have both the links of infinite channel capacity, all of the incoming traffic is routed onto the *faster* link for the entire duration of network operation. We then consider the specific case where only one unit has a link of finite channel capacity. Under the assumption that the network operation starts with an initial buffer occupancy in this unit which is less than the saturation value (of the buffer associated with the link of finite channel capacity), we derive the set of equations required to specify the optimal routing strategy. Since the solution to this set of equations require the knowledge of the traffic pattern for the entire duration of network operation, and this in turn necessitates an off-line computation, an on-line implementable suboptimal algorithm is suggested along the lines as done for the individual unit in Chapter 3. Some numerical examples comparing the performance of the optimal and suboptimal strategies are given.

We then consider a network topology composed of the three-node structure of Chapter 4. In Chapter 4, we have specified the optimal routing strategy for the constituent unit in terms of a single switching instant, when all the links of this unit are of infinite channel capacity. We consider a situation in which all the units of this larger topology have links of infinite channel capacity. The *globally* optimal strategy for this network is *bang-bang* in nature and is independent of the input traffic. The performances of the network under the *globally* optimal strategy and under the strategy synthesised from *locally* optimal strategies for the constituent units are compared for various choice of network parameters in order to get a quantitative idea of the performance difference.

Finally we consider a four-node hub network in Chapter 5, which can be topologically viewed as being composed of the three-node units of Chapter 4. Under the assumption that all the links of this network have infinite channel capacity, we investigate the nature of optimal routing strategy in this network. It is shown that the optimal routing strategy for this network has certain interesting properties as in the case of the constituent unit. The network operation ends with a direct routing of packets at all the three source nodes. The optimal routing strategy for this

network also, is shown to have the *loop-free* property. The conditions on the link parameters under which the optimal strategy for this network is the same as that synthesised from the *locally optimal* strategies for the constituent units, is obtained. When these conditions are not satisfied, a suboptimal way of traffic routing in this network can be obtained from the *locally optimal* strategies for the network units. We compare the performance of this algorithm with that of the optimal one in order to get a quantitative idea of the difference (in performance). Finally in Chapter 6, we summarise the major results in this thesis and indicate some directions for future research.

References

1. K. R. Krishnan, "Adaptive State-Dependent Traffic Routing Using On-line Trunk Group Measurements", *Teletraffic and Data Traffic in a period of Change*, ITC-13,1991.

2. L. Kleinrock, *Queueing Systems Vol I : Theory*, Wiley Interscience, 1975.

3. Janusz Filipiak, *Modelling and Control of Dynamic Flows in Communication Networks*, Springer-Verlag, 1988.

4. T.B. Crabill, D. Gross, M.J. Magazine, "A Classified Bibliography of Research on Optimal Design and Control of Queues", *Operations Research*, vol.25, No.2, March-April 1977.

Acknowledgements

A doctoral programme being the culmination of my academic dream, the time of submission of the thesis calls for a recapitulation of the chain of events that has led to it. Among the many things that flash before my mind, I recall the starting point of this long venture. For someone like me who was virtually cut off from academic activities for a period of four years since graduation, it was the inspiration and motivation of my close friend, Teju, which rekindled the desire to enroll for this programme. It was she who initiated the first step in this long chain of events by bringing to my attention an advertisement in the newspaper for the Ph.D interview at IIT Kanpur. She has stood by me in all my phases of ups and downs in the last seven years and I sincerely thank her for motivating me to take up this challenge, and for constantly supporting me in all the ways she possibly could, for the entire tenure of this programme.

I consider myself to be extremely fortunate to have Prof. P.R.K.Rao and Prof. S.K. Bose as my thesis supervisors. Prof. Rao has been a constant source of inspiration and has constantly encouraged me to strive harder. His willingness to listen to my presentations and his patient approach of rectifying the mistakes and suggesting improvisations have been the main factors that have helped shape this thesis. His constant desire for knowledge and his uncompromising stand to ward off complacency and mediocrity have been great eye openers for me. I am sure that whatever he has taught me both by way of his own example and by way of my frequent interactions with him, will always remain as ideals to constantly strive for in my life. Words are truly inadequate to express my gratitude towards him.

Dr. Bose has been constantly encouraging me to speed up the work. His presence has been one of an inspiring dynamism combined with a sharpness to steer clear of unwanted directions of investigation. He has also motivated me to look at the practical aspects of research without getting too much lost in the purely mathematical abstractions. I sincerely thank him for the concern for the progress of the thesis and for his constant moral support and encouragements during this

long tenure.

I recall with gratitude the pleasant associations with Dr. Srivathsan and the staff at Ernet Lab with whom I worked for a period of one year as a project employee. Dr. Srivathsan's concern for the progress of my thesis and his willingness to offer help can certainly not be repaid.

My friends Ashish Jadhav, Shafi, Chaveli, Balvinder, TSR, Mudgil, Sanjay and Atul Verma have helped me considerably in various stages of the thesis preparation despite their own busy schedules. I thank Ashish especially for helping me out with the preparation of the transparencies for the open seminar. Shafi has always been willing to help me out with LaTeX. I also thank my friends Venkatesh, Deepak Murthy, Kanetkar, Abhay Karandikar, Ajit Chaturvedi, Jayatheertha and Sandeep Agarwal, with whom I have shared the many joys and sorrows of doing the Ph.D.

Finally I thank my parents, my sister and her family who have stood by me and have supported me in every possible way within their reach, for the last seven long years. Unnikrishnettan and family have always provided me with a homely atmosphere away from the academic world and I sincerely thank them too.

Vinodkumar.

Contents

List of Tables

List of Figures

Glossary of Symbols

Mathematical Symbols	Explanation
$\rightarrow$	tending to
$\Rightarrow$	implies
$\forall$	for all
$\in$	belonging to
$\equiv$	is identically equal to (over an interval)

Variables	Explanation
D_{lm}	Delay in the steady state on the link (l,m).
f_{lm}	Flow on link (l,m) in message/second.
C_{lm}	Channel capacity of link (l,m) in messages/second.
$x_{jk}(t)$	Buffer occupancy at time t in the buffer associated with link (j,k).
$\alpha_{jk}(t)$	Fraction of the total traffic arriving at node j, routed onto the link (j,k).

$\lambda_j(t)$	External traffic arriving at node j.
J	Performance Index
H	Hamiltonian
$p_{jk}(t)$	Costate variable
a_{jk}	Link parameter for the link (j,k).
x_{jks}	The value of the buffer occupancy at which the rate of flowout from link (j,k) reaches the maximum value C_{jk}.

Chapter 1

Introduction

1.1 The Problem of Routing : A Control-Theoretic Perspective

Considerable amount of attention has been devoted in the recent past, to an investigation of efficient traffic control rules for communication networks. Among the many aspects of traffic control, an important one which has been intensively studied is the problem of synthesising an efficient routing strategy. A routing strategy[1] provides a set of rules that specifies the flow of messages or packets through the network from the source to the destination. At any switch along the path, the input traffic is routed from its input line to one of the output lines according to this (specified) set of rules and the destination of packet. An improperly conceived routing strategy could cause inefficient utilisation of the network resources, unduly long delays, and sometimes loss of message integrity because of out-of-sequence and duplication of packets. The problem of obtaining efficient routing strategies for fast and reliable delivery of messages to their destination, is therefore of utmost importance in the operation of modern data and voice communication networks.

Based on how the routing strategies incorporate the variations in the traffic

[1]also termed as routing algorithm/procedure/policy in literature. In this thesis too, these terms are used interchangeably.

and/or the network conditions, they can be classified into the following three categories[2]

- *static routing strategies,*
- *quasi-static routing strategies,*
- *dynamic routing strategies.*

In a purely static situation, given fractions of the traffic at node i of the network destined for each of the other nodes $j \neq i$ are directed on each of the links outgoing from node i. These fractions are decided upon before the network starts operation, and are fixed in time and depend only on the time and ensemble averages of the message flow requirement.

In the case of quasi-static routing, changes of routes are allowed only at given intervals of time and/or whenever extreme situations (eg. link failure) occur. The time intervals between routing changes will usually be relatively long, so that most of the time messages will be delivered in order and will not need individual addressing, but on the other hand, if a link fails or recovers or if the traffic and delays build up in a particular section of the network, the routing will be changed accordingly.

The completely dynamic strategy, on the other extreme, allows continuous changing of routes depending upon the instantaneous system states and traffic conditions.

One of the best known analytical model for static routing in data communication networks is that proposed by Kleinrock in 1962 [14]. The salient features of this routing strategy are the following:

In [14], it was assumed that messages arrive from outside the network to the nodes according to independent constant rate Poisson process, and their lengths were assumed to be independent exponentially distributed random variables, and

[2]Although this classification is extensively found in the literature, the distinction between these, is somewhat artificial. It is often difficult to draw precise boundaries between them.

independent of the arrival times. It was assumed[3] that the service time of a message is chosen independently at each node in the analysis. Based on an M/M/1 queue analysis, the delay in the steady state in each link is calculated explicitly as

$$D_{lm} = \frac{f_{lm}}{(C_{lm} - f_{lm})}$$

where

f_{lm} is the flow in link (l, m) in messages/second.

C_{lm} is the capacity of link (l, m) in messages/second.

D_{lm} is the total delay/unit time experienced by all the messages in link (l, m).

The routing procedure is so obtained as to minimise the total delay over the entire network $D_t = D_{lm}$.

While static routing algorithms such as the one suggested in [14] and various other types are still employed in many networks today, and are relatively simple to implement, changing situations in real networks, such as a line failure or a change in the traffic distribution, necessitates some degree of adaptivity. The adaptive routing algorithms must perform the following functions [1]:

(1) Measurement of the network parameters pertinent to the routing strategy.

(2) Forwarding of the measured information to the point(s) (Network Control Centre (NCC) or nodes) at which routing computation takes place.

(3) Computation of routing tables.

(4) Conversion of routing table information to packet routing decisions.

Dynamic routing procedures which incorporate the above adaptive features, are known to offer significant improvement in the efficiency provided a proper exploitation of the regular and random traffic variations is carried out. While the additional computational overheads for the implementation of the above mentioned functions are high, the emergence of faster computing machines and the

[3]known as the Kleinrock's Independence assumption.

continuing improvements in computer technology, has reduced this problem to a large extent. Therefore there is a shift in trend from static routing algorithms to dynamic routing algorithms in the recent times [9,10,11]. Many of the computer communication networks such as ARPANET, TELENET and TYMNET currently use dynamic routing algorithms to compensate for input traffic variations and to respond to changes in topology, and to take advantage of temporary changes in loading in different paths. Such a trend (from static to dynamic routing) is perceived in the case of circuit-switched networks also. State-dependent routing schemes for telephone traffic, in which the route for a call is chosen on the basis of the occupancy-state of the trunk groups at call arrival, have received much attention in the last few years [16,17,18,19]. An excellent survey of the various dynamic routing strategies employed in circuit-switched networks is given in [15].

Given the above scenario, let us look at the theoretical frameworks within which the problem of routing, and more specifically that of dynamic routing, has been investigated. Although there is an abundance of literature available on all the three categories of routing algorithms [that have been] implemented in the existing networks, [1,2,3,4,5], it appears that there is a paucity of basic models and theories which are capable of providing deeper analytical insights into them. Such a lack of analytical insight renders the task of a quantitative comparison of the performances of these algorithms very difficult. Intuition, heuristics, and simulation, rather than a detailed and rigorous analysis, have very often been the basis for the selection of a routing algorithm in many of the existing networks. While traditionally traffic phenomena in communication networks have been analysed with queueing models [6,7], these traditional analytical techniques are usually inadequate in dealing with networks with state sensitive dynamics. Classical queueing models, being too microscopic in their description, very often lead to overwhelmingly complex system models. Besides, under non-stationary environments queueing analysis is very often found to be intractable. Perhaps the most serious limitation to the use of queueing models in the context of routing is that, these descriptive queueing models tell us something at best about the operating characteristics of an existing (or

proposed) system, operated under a specified control policy [eg: routing strategy]. The task of synthesising the best (or at least a good) control policy for a system is seldom addressed to.

An increasing number of alternative models some of which at least, are prescriptive [50] in nature (as opposed to descriptive queueing models) have been proposed in the context of traffic control (flow control and routing) of networks in the recent times. The following studies reported in the literature illustrate this:

- T.Yum and M.Schwartz [9] considered the approach of superimposing a locally adaptive algorithm (called join-biased-queue (JBQ) rule) on a fixed traffic flow distribution obtained from the best stochastic (BS) rule suggested in [20]. The resultant JBQ-BS rule was analysed on small networks and has been shown to provide certain delay improvement over the BS rule.

- G.Foschini and J.Salz [10], using the framework of diffusion theory, have analysed a dynamic routing strategy where messages arriving at a certain node are routed to leave the node on the link having shorter queue.

- F.H.Moss and A.Segall [11] have investigated the application of optimal control theory to the problem of dynamic routing in networks.

- A learning automata based approach to routing in telephone networks has been developed by Narendra and Thathachar [8]. Chrystall and Mars have investigated the use of learning automata in the case of message switched traffic.

- The framework of Markovian decision theory has been used to describe the routing of calls in circuit-switched networks. An excellent introduction to this is given in [29].

Thus, a growing need to incorporate adaptivity (to the changes in the network conditions and/or the traffic pattern) and the need to incorporate a **synthesis approach** rather than an **analysis approach** have resulted in the introduction of

estimation and optimal control methods also in the design of flow control and routing algorithms. A significant contribution in this direction in terms of providing a systematic approach to the synthesis of advanced traffic control rules (within the framework of optimal control theory) was made by J.Filipiak [21]. The approach adopted by him in [21] consists in the modelling of dynamic flows. The basic model of dynamic flows relates the growth in the amount of packets/messages in the system at time t, by means of a deterministic differential equation, to the following three quantities:

(i) The intensity of newly arriving traffic

(ii) The intensity of traffic discarded/rejected by the system and

(iii) The intensity of successfully delivered traffic.

Expressed mathematically,

$$\dot{x}(t) = -[\lambda^c(t) + \lambda^r(t)] + \lambda^{in}(t) \tag{1.1}$$

where

$x(t)$= amount of traffic in the system at time t

$\lambda^{in}(t)$=intensity of newly arriving traffic

$\lambda^r(t)$=intensity of traffic discarded or rejected by the system, and

$\lambda^c(t)$= intensity of successfully delivered traffic.

It was assumed by Filipiak [21] that the intensity of the outgoing traffic $\lambda^c(t) + \lambda^r(t)$ can be approximated by a non-linear function $\mu G(x)$ of the system state $x(t)$. The intensity of arriving traffic $\lambda^{in}(t)$ is the fraction $\alpha(t)$ of the total traffic arriving at the node from which the link originates.

Thus the state of the network at time t, $\underline{X}(t)$ can be given by the differential equation:

$$\dot{\underline{X}} = \underline{f}(\underline{X}, \underline{\alpha}, \underline{\lambda}(t)) \tag{1.2}$$

In the Equation (1.2) above,

$$\underline{X} = \begin{bmatrix} \vdots \\ x_{jk}(t) \\ \vdots \end{bmatrix}$$

where $x_{jk}(t)$ denotes the number of packets in the buffer associated with link (jk) at time t,

$$\underline{\alpha} = \begin{bmatrix} \vdots \\ \alpha_{jk}(t) \\ \vdots \end{bmatrix}$$

where $\alpha_{jk}(t)$ denotes the fraction of the total traffic arriving at node j, routed onto link (jk),
and

$$\underline{\lambda} = \begin{bmatrix} \lambda_1(t) \\ \lambda_2(t) \\ \vdots \\ \lambda_n(t) \end{bmatrix}$$

where $\lambda_i(t)$ is the total external traffic arriving at node i.

The problem of optimal routing (dynamic routing) is then formulated as finding the $\underline{\alpha}(t)$ such that a suitably chosen performance functional is optimised. A detailed description of this approach is given in Chapter 2 of this thesis.

1.2 Motivation for this Thesis

As mentioned in the previous section, the possibility of treating the network as a dynamical system within the dynamic flow model formulation allows us to apply the powerful apparatus of optimal control theory for the synthesis of optimal traffic routing strategies. Using a familiar technique of optimal control theory, namely Pontryagin's Maximum principle, the necessary conditions the optimal state, costate and the control (routing) variables have to satisfy can be obtained.

These necessary conditions are given by a two point boundary value problem in the set of state and costate variables. It is well known that it is not easy to obtain the solution to these equations because the transversality conditions (boundary conditions) fix the values of the costate variables at the end of the time interval. The problem becomes more complicated when the dimensions of the network is large, since the dimensionality of the system of differential equations in the state and costate variables also increases correspondingly. The emphasis in Filipiak's work [21] is to consider the steady-state solution to the costate variables when the duration of the network operation T tends to infinity. Under this assumption, the solution to the costate variables (steady state) can be obtained (and from which the optimal routing variables) by solving a set of algebraic equations.

In this thesis, we relax this assumption of an infinite time duration operation of the network. Since in practice, networks are subjected to shut downs and need rebooting from time to time, we consider that the assumption of a finite time duration of operation corresponds to a practically more realistic situation. Furthermore, our analysis doesn't impose any restrictions on the value of T, and can therefore be viewed theoretically as a more general case (T equals infinity being a special case).

As mentioned in the previous section, the flow out from any buffer onto the corresponding link was assumed to be a non-linear function $\mu G(x)$ of the buffer occupancy in [21]. In a packet switching context it is reasonable to assume that the flow out function increases with increasing buffer occupancy and saturates at a value equal to the channel capacity of the link. It is also reasonable to assume that the flow out on a link is zero when the buffer is empty. Under these set of assumptions, a good approximation to the flow out function is one in which it increases linearly with the increase in buffer occupancy and saturates at a value equal to the channel capacity of the link. We consider that this class of flow out functions (in which they increase linearly with increase in buffer occupancy and saturate at the channel capacity) as against more complex form of functions for the flow out given in [21], will make an interesting study in itself. The emphasis

in our work is on this model. A different context in which such a model best describes a practical situation, is one in which the network is operating under a steady state (under a static routing strategy) and the sudden traffic fluctuations (perturbations) are to be routed (dynamically) in the network so as to achieve optimality. A linearized flow model, we consider, will best describe this practical situation.

The performance functional which is chosen (in this thesis) to be optimised is the total buffer occupancy time. The rationale behind choosing this is that a waiting cost is incurred at a rate proportional to the number of customers in the system. Hence whenever we use the term *optimal routing strategy* in this thesis, we mean a routing strategy which minimises the total buffer occupancy time for the entire duration of network operation.

Thus having chosen the flow model and the performance criterion to be optimised, we approach the problem of synthesising the optimal routing strategy for a large network in the following manner.

- Identify simpler structures within the large network.
- Obtain the optimal (or at least a good suboptimal) routing strategy for these simpler structures.
- Synthesise the routing strategy for the large network from the locally optimal/suboptimal routing strategies.

This thesis is motivated by the need to study the nature of optimal routing strategy and to specify it, in simple network units which are constituents of larger networks. We identify two such simple structures, namely a two-node structure (given in Chapter 3), and a three-node structure (given in Chapter 4). The problem of optimal routing in these network units is investigated in detail and the optimal routing strategy is specified under certain practically realistic assumptions. When the specification of the optimal strategy requires off-line computation, we suggest an on-line implementable suboptimal strategy for the network units. We also

look at some network topologies which are composed of these network units and study the properties of the optimal routing in these networks in Chapter 5. The main contributions from our investigations and the organisation of this thesis is summarised in the following section.

1.3 Contributions and Organisation of the Thesis

In Chapter 2, we review the various modelling schemes used in the performance study of networks. We review the dynamic flow model proposed by Filipiak [21] in detail and highlight the reasons for the choice of this modelling scheme in the synthesis of traffic control rules. A brief survey of the routing strategies employed both in the context of circuit switched networks and packet switched networks is also presented in this chapter.

In Chapter 3, we investigate the problem of optimal routing in a two node network, in which the *faster* link has a finite channel capacity. Under the assumption that the initial buffer occupancy on the *faster* link is below the saturation value (we argue that this assumption, though may appear to be overly restrictive, is not quite so in practice), we derive the set of equations in terms of the link parameters and the input traffic, the solution to which specifies the optimal routing strategy. Solution to these equations requires the knowledge of the load pattern for the entire duration of network operation, and this in turn necessitates an off-line computation. We therefore propose an on-line implementable suboptimal routing strategy for this network. The performance of the optimal and the suboptimal strategies are compared in the case of some numerical examples.

In Chapter 4, we investigate the problem of optimal routing in a three node network. We start with the assumption that all the links of the network have infinite channel capacities. The implication of this assumption on the network operation is that packets can be sent out on the links at a rate proportional to the buffer occupancy. The optimal routing strategy in this case is *bang-bang* and can be completely specified in terms of a single switching instant. The equation to

specify this switching instant is also obtained in terms of the link parameters of the network and the duration (T) of operation of the network. It is shown that the optimal routing strategy has the *loop-free* property.

We then relax the assumption of infinite channel capacities and consider the case wherein one of the direct links has finite channel capacity. We show that the *loop-free* property holds good under this situation also. The routing strategy need not be *bang-bang* in nature (we provide examples wherein the routing variables take non-zero, non-unity values over intervals). The various modes of network operation are defined and it is shown that out of the 27 possible modes, only 4 are valid in a terminal interval. These cases are illustrated with examples.

We then impose certain additional assumptions on the link parameters of the network, which in practical terms imply that corresponding to a situation in which all the links have infinite channel capacities, the optimal routing strategy is to route the entire traffic arriving at a source node onto the direct link for the entire duration of network operation. Under the assumption that the initial buffer occupancy on the link with finite channel capacity is below the saturation value, the set of equations (in terms of the link parameters and the input traffic) required to specifies the optimal routing strategy is derived. Solution to these equations requires the knowledge of the load pattern for the entire duration of network operation and this in turn necessitates an off-line computation of the routing strategy. As in the case of Chapter 3, we propose an on-line implementable suboptimal strategy for this network. The performances of the optimal and suboptimal strategies are compared in the case of some numerical examples.

In the final section of Chapter 4, we investigate the case wherein the flow out on a link is an exponential function of the (associated) buffer occupancy. The optimal routing strategy in this case is obtained by numerically solving the two point boundary value problem in the state and costate variables. From the numerical investigations carried out for various choice of link parameters, initial buffer occupancies and input traffic, we conjecture that the optimal routing strategy has the *loop-free* property. It is also observed that there is at most one switching instant and

that the network operation always ends with a direct routing of packets at both the source nodes.

In Chapter 5, we investigate the problem of optimal routing in some networks which can topologically be viewed as being composed of the two node network units and the three node network units. We first consider a network which is composed of the two node units composed in Chapter 3. For the case wherein m such network units (of the total n which constitute the topology) have their *faster* link of finite channel capacity (both the links of the remaining $(n-m)$ units are assumed to be of infinite channel capacity) we study the properties of the optimal routing strategy. It is proved that in all the network units which have both the links of infinite channel capacity, all of the incoming traffic is routed onto the *faster* link for the entire duration of network operation. We then consider the specific case where only one unit has a link of finite channel capacity. Under the assumption that the network operation starts with an initial buffer occupancy in this unit which is less than the saturation value (of the buffer associated with the link of finite channel capacity), we derive the set of equations required to specify the optimal routing strategy. Since the solution to this set of equations require the knowledge of the traffic pattern for the entire duration of network operation, and this in turn necessitates an off-line computation, an on-line implementable suboptimal algorithm is suggested along the lines as done for the individual unit in Chapter 3. The performances of the optimal routing strategy and of the suboptimal strategy are compared in the case of some numerical examples.

We then consider a network topology composed of the three-node structure of Chapter 4. In Chapter 4, we have specified the optimal routing strategy for the constituent unit in terms of a single switching instant, when all the links of this unit are of infinite channel capacity. We consider a situation in which all the units of this larger topology have links of infinite channel capacity. The *globally* optimal strategy for this network is *bang-bang* in nature and is independent of the input traffic. The performances of the network under the *globally* optimal strategy and under the strategy synthesised from *locally* optimal strategies for the constituent

units are compared for various choice of network parameters in order to get a quantitative idea of the performance difference.

Finally we consider a four-node hub network in Chapter 5, which can be topologically viewed as being composed of the three-node units of Chapter 4. Under the assumption that all the links of this network have infinite channel capacity, we investigate the nature of optimal routing strategy in this network. It is shown that the optimal routing strategy for this network has certain interesting properties as in the case of the constituent unit. The network operation ends with a direct routing of packets at all the three source nodes. The optimal routing strategy is shown to have the *loop-free* property also. The conditions on the link parameters under which the optimal strategy for this network is the same as that synthesised from the *locally optimal* strategies for the constituent units, is obtained. When these conditions are not satisfied, a suboptimal way of traffic routing in this network can be obtained from the *locally optimal* strategies for the network units. We compare the performance of this algorithm with that of the optimal one in order to get a quantitative idea of the difference (in performance). Finally in Chapter 6, we summarise the major results in this thesis and indicate some directions for future research.

Chapter 2

Network Models and Routing Strategies: An Overview

Facilities for collective use are often encountered in a variety of situations in the present day society. Ticket windows, beds in a hospital, road junctions, communication trunks in a telephone network, the central processing unit in a computer are some of the frequently encountered examples. These facilities are usually of limited capacities, and therefore are prone to phenomena like congestion and loss when the demands of the users surpass the capacity. For an efficient (and economical) usage, it is essential to evaluate the performance quantitatively and to incorporate methodologies/strategies which enhance the performance. In the specific area of telecommunication systems, it is required to evaluate the Grades of Service (GOS) quantitatively, and to clarify relations between the GOS and the amount as well as configuration of the facilities. This, in turn, necessitates the use of appropriate models for the performance analysis of such facilities.

We mentioned in Chapter 1, that a number of alternative models for traffic and network characterisation have emerged in the recent times. Quite apart the extensive use of traditional techniques like simulation and queueing analysis, these models have also been employed (and perhaps, even substituted the use of traditional techniques) in the context of network management. We now review some

of the important models employed in the performance study of communication networks in the first section of this chapter. It is not our claim that what is given here is an exhaustive survey of all the models that have been suggested in the literature. Such an exhaustive survey is beyond the scope of this thesis. The presentation has been made here with a view to highlight the important ones that have been suggested in this context. The dynamic flow model, which is the framework of our study in this thesis, is explained in detail. We also take a cursory review of routing strategies in terms of some of the classifications which have frequently been followed in the literature and finally establish the reasons for the choice of flow models in the context of routing.

Broadly speaking, the modelling schemes that are employed in the performance evaluation of telecommunication networks can be classified into two major classes, namely, the simulation models and the analytical models.

2.1 Simulation Models

Simulation is defined [41] as a numerical technique for conducting experiments on a digital computer, that involves mathematical and logical models that describe the behaviour of the system (or some components thereof) over periods of time. The major reason for conducting this experiment on a computer is that it permits the study of real system without modifying the system in any way. Due to the increasing speed and decreasing cost of electronic computers, as well as development of programming languages particularly suitable for simulations, there has been a dramatic increase in the use of computer simulation in the recent years. In the context of computer communication networks, simulation is used at various stages of design and operation. Some of the main applications of simulation in the context of networks are the following :

- Computer simulation is applied as an alternative approach to cases where exact analytical solutions are not available. It is often difficult to arrive at exact

analytical solutions in the performance analysis of large-scale communication networks with complex congestion and/or routing control mechanisms.

- Simulation is used not only for cases wherein analytical solution is not obtained, but also in the cases where numerical computation of analytical solution is difficult. It is also used for validating the accuracy of approximate solutions.
- Simulation is used in the various stages of design and operation of networks, such as evaluating the performance to confirm the system application range, estimating the response to fault and over-load to determine the counter measures etc.

Simulations can be broadly classified into two:

- Continuous simulation
- Discrete simulation

Traffic simulation in the context of communication networks, is usually discrete-event simulation. The discrete state (number of calls/packets) of the system is changed by an event of call origination/packet arrival or termination/packet departure. As shown in Figure 2.1, there are the following three viewpoints [48] for modelling in a discrete-event simulation: *event*, *process* and *activity*.

(1) Event-oriented Modelling:

A simulation model is formulated by describing state changes by events such as packet arrival or departure.

(2) Process-oriented Modelling:

This modelling describes the behaviour of the entity (call/packet) in the system.

(3) Activity-oriented Modelling:

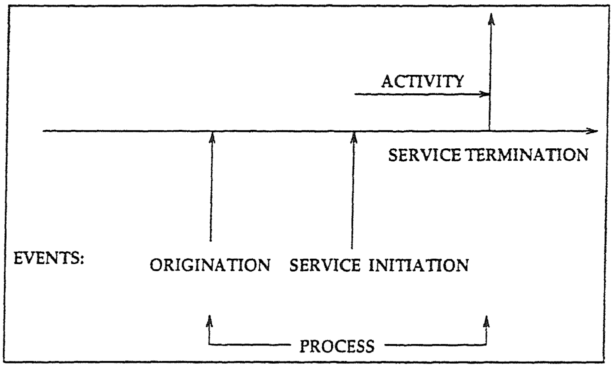

Figure 2.1: Viewpoints in Simulation

This modelling describes the time instants of initiation and termination of activities such as call duration and trunk busy. This method is suitable for modelling systems for which holding time depends on the state of the system. Since the simulation is performed by scanning the activities, the execution time is usually longer than for event-oriented modelling.

Programming for simulation may be done either in *general purpose languages* or in *simulation languages*. The main functions of the simulation languages are generation of random numbers, execution of time schedules and data saving, statistical analysis and output format conversion.

Although simulation is a powerful technique and is quite extensively used, there are certain disadvantages with it. Simulation is often computationally expensive and is a time-consuming process (especially if reliable statistics is to be obtained) and often not suited for real-time applications. Moreover, the proper design of a simulation experiment and the accurate interpretation of the output data are themselves difficult tasks. Although several new approaches have been developed in the recent years, aiming at improving the efficiency of simulation (in the sense of extracting as much information as possible from a single run), much work has yet to be done to eliminate the above inadequacies.

2.2 Analytical Models

An analytical model is a representation of the system that specifies a functional relationship between the system parameters and a chosen performance criterion in terms of mathematical equations. In the context of communication networks, the earliest attempt to use analytical models to predict the performance began with the pioneering work by the Danish mathematician A.K. Erlang[1] on telephone switching systems [26,27]. In the voice telephone network, demands for service take the form of subscribers initiating a call. Erlang found that given a sufficiently large population, the random rate of such calls can be described by a Poisson process. An important and perhaps most extensively used tool for analytical modelling, namely queueing theory, has its foundations in this work by Erlang. The 1920's were basically devoted to the application of his results and it was not until the mid-1930s, when Feller [28] introduced the concept of the birth-death process that queueing theory was recognised by the world of mathematics as an object of serious interest. The later developments in queueing theory occurred in the 1940's and 1950's from the contributions made by C. Palm, F. Pollaczek, A. Ya Khinchin and others. Much mathematical work stemmed from an important paper by D.G.Kendall [34] who introduced the classification of queues by input, service mechanism and queue discipline.

Descriptive queueing models for communication networks are most often based on the theory of Markov processes, either in discrete time or in continuous time. Markov-process theory makes it possible to set up a system of equations for the equilibrium or time-dependent probabilities for the system state (e.g. number of customer/packets present). It is sometimes possible to solve these equations analytically, often with the aid of transforms as in the case of simple systems such as M/M/1 queues. The result may be closed-form expressions for measures of effectiveness such as the average number of packets/customers present or the

[1]Historically this work by Erlang was done even before probability theory was popular or even well developed.

mean waiting time of a customer.

Incorporating the effects of stochastic (random) variation in a model of queueing system can easily result in mathematical intractability, especially if the real system is more complicated than the one referred to above. Unfortunately, virtually all real-world queueing systems are more complicated. There may be more than one server or a finite number of places in the buffer. Jobs may balk (refuse to enter the system) or renege (leave after entering the system). Different jobs may have different priorities and be processed accordingly, rather than the order of arrival. Jobs may arrive in batches and/or be served in bulk. The server may be turned on or off, due to random breakdowns or the need to perform other functions. The pattern of job arrivals may exhibit predictable long-run variations (e.g. slack versus peak periods), in addition to the short-run, stochastic variations.

Besides the above difficulties, there are other limitations to queueing analysis, especially in the context of analysis and synthesis of traffic control rules. Some of these are [21],

- Transient solutions can be obtained for simple systems only.
- The dimensionality of the network model is usually high.
- Detailed knowledge of time-dependent probabilities is rarely available.
- Queueing analysis is often difficult for nonstationary traffic and network conditions.

Other teletraffic models, which aim to eliminate, at least some of the above difficulties associated with queueing models, have been suggested in the literature. The diffusion approximation given below, is sometimes used to avoid the difficulty of obtaining explicit transient solutions. We look at this model in detail in the following subsection.

2.2.1 Diffusion Models

Diffusion theory has sometimes been successful in providing excellent approximate solutions to difficult queueing problems. The approach in diffusion approximation is to replace a discrete process like the vector of queue lengths, Q(t), with a continuous path Markov process with drift, q(t), i.e. a diffusion. The diffusion process q(t) has a dynamical description via a stochastic differential equation

$$dq(t) = a(q)dt + D^{1/2}(q)dw(t) \tag{2.1}$$

where the vector $a(q)$ is the differential mean, the matric $D^{1/2}(q)$ is the positive definite square root of the differential dispersion matrix and dw(t) is a standard zero mean white Gaussian noise process. That is

$$a(q) = \lim_{h\to 0} h^{-1}E(q(t+h) - q(t)) \tag{2.2}$$

$$\text{and } D(q) = (d_{ij}(q)) \tag{2.3}$$

$$= \lim_{h\to 0} h^{-1}E[(q_i(t+h) - q(t))(q_j(t+h) - q_j(t))] \tag{2.4}$$

A central result of diffusion theory is that the transition density for $q(t)$ satisfies the Fokker-Planck (FP) equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{k} \frac{\partial a_k(q)p}{\partial q_k} + 1/2\sum_{1}^{k}\sum_{1}^{k} \frac{\partial^2 d_{ij}(q)p}{\partial q_i \partial q_j}. \tag{2.5}$$

The equilibrium density, when it exists, satisfies the above equation with $\frac{\partial p}{\partial t} = 0$.

So the diffusion equilibrium density is obtained by solving a second order partial differential equation (PDE) while the desired equilibrium probabilities for the discrete vector queue satisfy a difference equation. The reason for turning to diffusion model is that it may be easier to solve a partial differential equation than a difference equation. Boundary conditions (BC) that, for example, demand that the quantities of interest such as queue lengths be positive are an essential aspect of the diffusion formulation. While the hope is that the resulting PDE and BCs will be more tractable than the analysis of the original system of difference equations, there is of course, no assurance that this will always be the case.

Diffusion theory has been used by several authors for the analysis of network behaviour [35,36,10]. An interesting application of this theory to the problem of dynamic routing was provided by G.Foschini and J.Salz in [10] for the case of a single node in a data network. The messages, or packets arriving at a certain node are routed to leave the node on the link having the shorter queue. For a node with K outgoing queues, under the assumption of heavy traffic, an optimum dynamic routing strategy (in the sense of minimising the average delay) was obtained.

2.2.2 PH-MRP Input Models

The Phase-Type Markov Renewal Process is usually used for representing versatile renewal and non-renewal processes appearing in modern teletraffic systems, such as the ATM for BISDN. A brief explanation of phase-type (PH) distribution is as follows:

In a continuous-time Markov chain with r transient states and a single $(r+1)^{st}$ absorbing state, suppose that upon entering the absorbing state, the process instantaneously jumps to the transient state j, $j = 1, 2, 3, ...r$, with probability α_j. The PH distribution is defined as the inter-visit time to the absorbing state, and is characterised by (α, T), where T is the transition rate matrix among the transient states, which is an irreducible r x r matrix. The row vector $\underline{\alpha}$ with component α_j is called the initial probability vector. The column vector T^o defined by

$$T^o = -Te$$

represents the transition rate from the transient states to the absorbing state, where e is the unit column vector with all components equal to 1. The PH distribution is said to be in *phase* j if the underlying Markov process is in state j. A renewal process with an interarrival time of PH distribution is called a *phase-type renewal process* (PH-RP).

If we modify the Markov chain above to have n absorbing states, with jumping probability α_{ij} from absorbing state i, $i = r+1, r+2, ..., r+n$, to transient state j,

$j = 1, 2, \ldots.., r$, then the successive visits to the absorbing states constitute the PH-MRP, in which inter-visit times follow PH distribution not identical in general and correlated with each other. The PH-MRP is said to be in phase j, if the underlying Markov chain is so. The PH-MRP is characterised by representation (α, T, T^o). The n x r matrix α with components α_{ij}, and the r x n matrix T^o are the extensions of the corresponding vectors above, and we have the relation,

$$T^o e = -Te$$

The arrival rate (averaging each PH distribution) is given by

$$\lambda = \pi T^o e$$

where π is the *stationary probability vector* of $T + T^o\alpha$ satisfying

$$\begin{aligned} \pi(T + T^o\alpha) &= 0, \\ \pi e &= 1. \end{aligned}$$

2.2.3 Traffic Model with Burstiness Constraints

Instead of using a stochastic model for entering traffic, Cruz [30,31] assumed that the entering traffic is *"unknown,"* but satisfies certain regularity constraints. The purpose of introducing this characterisation of the traffic was to facilitate the study of the following three parameters of the network [30]:

1. *Delay*

2. *Buffer Allocation Requirements*

3. *Throughput*

The constraints the traffic stream have to satisfy have the effect of limiting its burstiness hence they were called "burstiness constraints." These burstiness constraints of the traffic were formally defined in [30] as below:

Cruz first defined the rate function $R(t)$ of a traffic stream flowing on a communication link as follows: For any $y \geq x$, $\int_x^y R(t)dt$ is the amount of data from the stream that is transmitted on the link in the interval $[x, y]$. Thus, in general, $R(t)$ represents the instantaneous rate of traffic from the stream flowing on the link at time t. The traffic flowing at any point in the network can be characterised (partially) using the burstiness constraints in the following way: Let $R(t)$ denote the rate function of the traffic. Given $\sigma \geq 0$ and $\rho \geq 0$, we say that the traffic satisfies the burstiness constraints (σ, ρ) (and write $R \sim (\sigma, \rho)$) if and only if for all x, y satisfying $y \geq x$ the following relationship holds:

$$\int_x^y R \;\leq\; \sigma + \rho(y - x). \tag{2.6}$$

Thus the traffic stream is said to satisfy a burstiness constraint if the quantity of data from the stream contained in *any* interval of time is less than a value that depends on the length of the interval. A different interpretation of the burstiness constraints of the traffic is that given any positive number ρ, there exists a (possibly infinite) number σ_ρ such that if the traffic is fed to a server that works at rate ρ while there is work to be done, the size of the backlog will never be larger than σ_ρ. The constant of proportionality ρ in the above determines the upper bound to the long term average rate of traffic flow if such an average rate exists.

Note that if packets are L bits long, then if $R(t) \in \{0, C\}$, it is possible that $\int_t^{t+L/C} R = L$ for some t. Thus if $R \sim (\sigma, \rho)$, then ρ must satisfy

$$\sigma \;\geq\; L(1 - \rho/C). \tag{2.7}$$

Cruz extended the above notion to more general constraints on the traffic stream than in Equation (2.6). Specifically, if b is any function defined on the nonnegative reals and R is a nonnegative function such that

$$\int_x^y R \;\leq\; b(y - x), \tag{2.8}$$

for all $y \geq x$, we write $R \sim b$. The function b is assumed to be nondecreasing here.

Thus after characterising the traffic (partially), using the above manner, Cruz considered several network elements that can be used as building blocks to model

a variety of communication networks. Each type of network element was analysed by assuming that the traffic entering it satisfies the burstiness constraints. Under this assumption bounds were obtained on delay and buffering requirements for the network element, and the burstiness constraints that the traffic that exits the element were also derived. Communication network models consisting of the interconnection of network elements (which are analysed in isolation) operating under a fixed routing strategy were considered in [31].

The characterisation of the traffic in terms of burstiness constraints is particularly of interest in situations wherein the traffic to be handled by the network is highly bursty, for instance, by variable rate video compression schemes. The fact that the traffic can be characterised by simple parameters that can in principle be negotiated between users and the network, makes this model very appealing from the point of view of applications, especially in the design of high speed networks for bursty traffic.

2.2.4 $\in$-Quantile Function Models

In the emerging B-ISDN networks based on Asynchronous Transfer Mode, one of the primary issues is the characterisation of heterogeneous traffic issued by multimedia users. Performing statistical multiplexing requires a finer characterisation of the statistical behaviour of the source. In particular, there is a need for measures which give insight into the short-term (for bursts) and long-term (average) behaviour of the source.

The motivation behind the definition [37] of the time $\in$-quantile function, given below, was the need to distinguish between the long-term tendency (mean rate) and short term bursts at the peak rate of the source behaviour, since these parameters entail different constraints on the design of congestion control mechanisms. Another important fact is that the performance requirements of the emerging networks in terms of the loss requirements (typically of the order 10^{-9}) places emphasis on the tails of distributions which naturally lead to the consideration of quantile

measures. We briefly describe the notion of time quantile measures below.

The notion of time quantile function (and its properties) can best be illustrated in the following way. Consider a slotted-frame transmission system like SONET. Assume that the cells from the source arrive at discrete-times, i.e. during slots and the slot duration be denoted by θ. Let $\{N_k\}_{k \in Z}$ be a discrete-time point process and let $\in$ be any real number such that $0 < \in \leq 1$. (N_k denotes the number of arrivals in $(0, k\theta]$). Define the function $M_{\in}$ of k as follows:

$$M_{\in}(k) \;=\; min\{n : Pr(N_k > n) \leq \in\} \tag{2.9}$$

$M_{\in}(k)$ is called the time $\in$- quantile at time k of process $\{N_k\}$.

Stated alternatively, $M_{\in}(k)$ is a measure of the maximum number of arrivals in an interval of length $k\theta$ for a given probability $\in$. From the definition of $M_{\in}(k)$, the following properties can be easily inferred for the model when there can be at most one arrival in a given slot.

1. $max_{k>0} \frac{M_{\in}(k)}{k} \leq 1$.
   For the general discrete-time point process, $sup_k \frac{M_{\in}(k)}{k}$ is the maximum number of arrivals in a slot.

2. $\frac{M_{\in}(k)}{k}$ is a decreasing function of $\in$ (for any fixed k) and an increasing function of k (for fixed $\in$).

Mazumdar *et al* [37] showed that the average rate of arrivals λ is related to the time $\in$-quantile of the process $\{N_n\}$ (provided that N_n is ergodic) in the following manner:

$$\lim_{n \to \infty} \frac{M_{\in}(n)}{n} = \lambda, \text{ for any } 0 < \in < 1.$$

Thus the average rate of arrivals can be recovered from the time-quantile of the process. The second asymptotic property related to the $M_{\in}(n)$ function is that a suitably scaled version converges to the limit of the indices of the dispersion of counts and intervals (for the definitions of the terms *index of dispersion of counts* and *index of dispersion of intervals,* refer to [37]).

The two main utilities of the time $\in$-quantile function characterisation of the traffic were shown to be the following [37]:

- The time $\in$-quantile function provides us with a statistical measure to assess the comparative performance of traffic streams from various sources.
- It was shown in [37] that the $\in$-quantile function can be used in connection with queue dimensioning. The role of $\in$ as a parameter related to the blocking probabilities was brought out.

2.2.5 Learning Automata Based Network Models

Mathematical models of the network using a class of learning algorithms, and studies related to the convergence and asymptotic performance of these algorithms have been discussed by several authors [38,39] in the context of traffic routing in circuit-switched networks. These models are particularly useful when little is known about the response of the environment to actions of the system. A formal description of an automaton is as follows:

An automaton is defined by the quintuple $\{X, \phi, \alpha, T, G\}$ where,

$$\begin{aligned}
X &= \text{The set of inputs to the automaton.} \\
\phi &= \text{The set of internal states.} \\
\alpha &= \text{The set of actions } \{\ \alpha_1, \alpha_2, \ldots.., \alpha_n\}. \\
T &= \text{The state-transition map at stage } n: \\
\phi(n+1) &= T[\phi(n), \alpha(n), X(n)]. \\
G &= \text{The action-selection function :} \\
G &: \ \phi \rightarrow \alpha.
\end{aligned}$$

The automaton is deterministic or stochastic depending whether T and G are deterministic or stochastic. The class of automata considered for routing, called *variable structure automata* are of the form

$$\phi \ = \ p_1, p_2, p_3, \ldots.., p_r$$

$$\sum p_j = 1, \ p_i \geq 0.$$
$$G(\phi) = \alpha_i \ \text{ with probability } p_i$$

The states are defined by a probability vector, and the action map is the selection of an action i with probability p_i.

The environment in which the automaton operates is characterised by a set of inputs $\alpha(n)$ and random outputs $x_i(n) \in X$. Of special interest is the case where the environment output is binary, that is, a reward ($x = 0$) or a penalty ($x = 1$). The operation of the automaton is as follows: The automaton outputs are fed into the environment, which reacts to these outputs by providing a reward or penalty. This response is fed back as input to the automaton, which provides another action, and so on. A learning behaviour is made possible by this feedback organisation: The automaton is provided with some performance function that, it is hoped, will eventually be minimised. The behaviour of the automaton is determined by the choice of the maps T and G, which in turn depend on the particular performance criterion that one is trying to minimise. The actual distributions for the response of the environment are unknown and the operation of the automaton in no way assumes any knowledge about these responses.

The application of this general automaton formalism to the routing of calls in circuit-switched networks is in the following way. Assume that in each node k of the network, there is an automaton, denoted $A_k^{i,j}$, for each traffic stream i, j going through the node. If the origin of the call is not available, the automaton is denoted A_k^j. Two sets of actions are possible for the automaton: (1) the set of all links out of the node and (2) the set of all *sequences* of links. The environment is the network beyond node k, and the responses are of the reward-penalty type, corresponding to a connected or lost call respectively. The objective, of course, is to minimise the number of lost calls. Various probability-update algorithms have been proposed in the literature [38,39]. For the purpose of illustration of the theory of learning automata in the context of routing, we give below the following algorithm:

The M-automaton suggested in [38] uses the choice of outgoing links as the

action set. Define $g_i(n)$ as the estimated probability of loss if action i is used at stage n, where action i is to route the call via group i. These state variables are updated by the

$$g_i(n+1) = g_i(n) + w[x(n+1) - g_i(n)]$$

where w is a constant and $x(n)$ is the network response at stage n and is given by

$$x(n) = \begin{cases} 0 & \text{if the call is completed} \\ 1 & \text{if the call is blocked.} \end{cases}$$

The action at each stage is to compute the current value of g_i, rank the groups in decreasing order of g, and attempt to route the call in this order.

More general algorithms, in which the update formula takes into account both reward and penalty (known as L_{R-P} schemes) have been suggested in the literature. An excellent introduction to this topic is given in [29].

The use of learning automata for routing has many advantages. The most obvious is the simplicity, both computationally and in terms of measurements. Learning automata assume nothing about the network or the traffic flows, and in particular do not assume that the arrival process of new calls is stationary. They are particularly well-suited for networks in which there are large variations in traffic. On the negative side are such questions as the stability of the algorithm, the rate of convergence, and the global optimality. Because of the difficulties associated with these questions, much research needs to be done before satisfactory answers can be given.

2.2.6 Dynamic Flow Models

Traffic routing and related areas of network management are currently undergoing review in a number of networks all over the world. The importance of reliable communication and the associated economic considerations account for the great interest in these problems in the past as well as the present. One of the main concerns of an efficient network management is the synthesis of an optimal routing

strategy. We argued in Chapter 1 that, from the perspective of **synthesising** control policies (e.g. routing strategies) in which adaptivity to network and/or traffic conditions can be incorporated, there is a need for a prescriptive model (as opposed to descriptive queueing models) of the system. We also gave a brief introduction to the flow model developed by Filipiak [21] in Chapter 1, and argued that the framework of flow model allows the network to be treated as a dynamical system and this, in turn, allows us to formulate the problem of optimal routing as an optimal control problem. We now give below a detailed description of the model suggested by Filipiak.

Let $A(t)$ represent the cumulative number of arrivals to time t, $t \in (0, T)$, $D(t)$ be the cumulative number of entities serviced during time $(0, t)$, $t < T$, and $Q(t)$ the number of entities in the system. From the conservation principle we get the relationship

$$Q(t) = A(t) - D(t) + Q(0). \tag{2.10}$$

Assume that the time behaviours of $A(t)$, $D(t)$ and $Q(t)$ have been measured on several occasions and the averages $\overline{A}(t)$, $\overline{D}(t)$ and $\overline{Q}(t)$ are known. Let us assume that these averages are continuous functions of time, differentiable piecewise in $(0, T)$. Then

$$\overline{Q}(t) = \overline{A}(t) - \overline{D}(t) + \overline{Q}(0) \tag{2.11}$$

Differentiating the above we get

$$\frac{d\overline{Q}(t)}{dt} = \frac{d\overline{A}(t)}{dt} - \frac{d\overline{D}(t)}{dt} \tag{2.12}$$

with the initial condition $\overline{Q}(0) = Q_0$.
Let $\lambda^{in}(t)$ denotes the average rate of arrivals, i.e. $\lambda^{in}(t) = \frac{d\overline{A}(t)}{dt}$, $x(t)$ denotes the average quantity in the system, i.e., $x(t) = \overline{Q}(t)$ and $\lambda^{out}(t)$ denote the average number of departures, we then get

$$\frac{dx(t)}{dt} = \lambda^{in}(t) - \lambda^{out}(t) \tag{2.13}$$

with the initial condition $x(0) = x_0$. It was assumed by Filipiak in [21], that if the system is not empty, then the intensity $\lambda^{out}(t)$ of the outgoing flow can be closely approximated by a function $G(.)$ of the system state $x(t)$, i.e.,

$$\lambda^{out}(t) = \mu G[x(t)] \tag{2.14}$$

where μ is the server capacity defined as the number of entities which can be served per unit time. Thus we find that evolution of the mean number in the system can be approximated by the following nonlinear differential equation

$$\frac{dx(t)}{dt} = -\mu G[x(t)] + \lambda^{in}(t) \tag{2.15}$$

with the initial condition $x(0) = x_0$.

The following interpretation was suggested in [21] for the function $G(.)$. For a constant input traffic $\lambda(t) = \lambda_0$, the system will asymptotically approach the steady-state $\overline{x}$ given by

$$\dot{x}(t) = 0 = -\mu G(\overline{x}) + \lambda_0 \tag{2.16}$$

Thus $G(\overline{x})$ is equal to ρ_0 and consequently, if the description given by Equation (2.15) is to be accurate for the steady-state, $G(.)$ must represent the steady-state utilisation factor as a function of the mean steady-state number in the system.

An alternative interpretation provided by Filipiak and several other authors [32,33] for the function $G(.)$ is that $\mu G(x)$ is a good approximation for the function $(1 - P_0(t))$ where $P_0(t)$ is the probability that the system is empty at time t. This can best be illustrated by considering an M/M/1 queueing system, in which the service times have an exponential distribution with constant parameter $\mu = 1$ and the input traffic is governed by a nonhomogeneous Poisson process with intensity $\lambda(t)$. The birth-death equations for this case can be written as

$$\frac{dP_0(t)}{dt} = -\lambda(t)P_0(t) + P_1(t) \tag{2.17}$$

$$\frac{dP_n(t)}{dt} = \lambda(t)P_{n-1}(t) - (\lambda(t)+1)P_n(t) + P_{n+1}(t) \tag{2.18}$$

Multiplying both sides of Equation (2.18) by n and summing from $n = 1$ to $n = \infty$, and making use of the fact that $\sum_{n=0}^{\infty} nP_n(t)$ is the average number in the system (i.e. $x(t)$), we get

$$\dot{x} \;=\; -(1 - P_0(t)) + \rho(t) \tag{2.19}$$

Comparing the Equation (2.19) above with Equation (2.15), it can be seen that $\mu G(.)$ represents the probability that the system is not empty at time t for an M/M/1 queueing system.

The first interpretation for the function $G(.)$ allows us to specify its functional form in the following manner. In the steady-state, we have the relation $G(\overline{x}) = \overline{\rho}^{in}$. Let us assume that the stationary averages $\overline{x}_n$ can be obtained from measurements, when the system is subject to load $\overline{\rho}_n^{in}$, $n = 1, 2, ..., N$. If the quality of the model is reflected in the smallness of the mean squared error, then to reduce the steady-state errors, we must minimise the expression $\sum_{n=1}^{N}(G(\overline{x}_n) - \overline{\rho}_n^{in})^2$. It is often convenient to choose a polynomial form for $G(.)$ and using the Gaussian technique, the polynomial coefficients which minimise the mean squared error can be obtained.

To summarise the above discussion, we see that flow models describe the time dependent average quantities of the network by means of deterministic ordinary differential equations. No assumption regarding the statistical properties of the input traffic is made in this formulation and hence the model can incorporate non-stationary traffic and network conditions.

2.3 Routing Strategies

We have so far considered some modelling schemes by which the performance of a network can be studied. Performance of a network depends on such factors as the network configuration, the offered load, and the network management methods. An important element of network management, called *network routing* consists of

the decision rules used to connect the calls as they arrive at the network; a variety of methods are now possible. The purpose of this section is to survey the routing techniques (also termed as algorithms/strategies) which have been suggested in the literature and to introduce classifications of the different schemes. These classifications often depend on the context in which the algorithm is studied. A complete classification of all the algorithms that have hitherto been proposed, is beyond the scope of this thesis. Moreover, there is no consensus on the precise meaning of many of the terms used to characterise routing. Therefore our approach below is to highlight some of the frequently used classifications (along with examples, wherever possible) that have been suggested in the literature. We look at them in two subsections, namely routing strategies for data networks and those for circuit-switched networks.

2.3.1 Routing Strategies in Data Networks

There are a number of ways to classify routing algorithms. A frequently used classification [40] is as follows:

(i) *Deterministic Algorithms*

(ii) *Stochastic Algorithms*

(i) **Deterministic Algorithms**

Algorithms in this class do not adapt to changes in traffic, but may be designed to provide satisfactory performance, on the average, over a range of traffic intensities. An example for this class is the *least-time delay algorithm*, wherein the routes for any source-destination pair are chosen to minimise the overall average time delay. *Shortest path algorithms* can also be considered to be belonging to this category. In *shortest path algorithms*, each communication link is assigned a positive number called its *length*. Each path (i.e. sequence of links) between two nodes has a length equal to the sum of the lengths of its links. A *shortest path routing algorithm* routes each packet along a minimum length between the origin and destination nodes of

the packet. Much work has been done on this [49] and more sophisticated versions, wherein the length of each link changes with respect to time and depends upon the prevailing congestion level of the link, have also been studied by various authors [49].

Another class of *deterministic algorithms* which is frequently referred to in the literature on routing is the *flooding algorithms*. In this type of routing, each node receiving the message simply retransmits it over all outgoing links, or a selected number of these following a simple rule. To limit the number of packet transmission, two rules are observed. First a node will not relay the packet back to the node from which the packet was obtained. Second, a node will transmit the packet to its neighbours at most once; this can be ensured by including on the packet the ID number of the origin node and a sequence number, which is incremented with each new packet issued by the origin node. By storing the highest sequence number received for each origin node, and by not relaying packets with sequence numbers that are less than or equal to the one stored, a node can avoid transmitting the same packet more than once on each of its incident links. This routing strategy is usually very simple and robust, but the network becomes flooded with multiple copies of any message. Therefore, this is appropriate under low traffic conditions only.

(ii) **Random (Stochastic) Algorithms**.

The simplest type of *random routing strategy* involves assigning fixed decision rules as to which neighbouring node to send the messages. The decision rule specifies the probability with which the messages at each node will be directed to the adjacent nodes. Algorithms of this type have been analysed by Kleinrock [14] and Proser [42]. These are simple from the implementation point of view, tend to be robust and are relatively insensitive to changes in network structure. However, they do not make use of any available information on the traffic patterns, desirability of certain routes, etc.

Another classification for routing algorithms is in terms of the adaptivity to the traffic and/or the network conditions. Accordingly routing algorithms are classified into the following three categories.

- *Static Routing Algorithms*
- *Quasistatic Routing Algorithms*
- *Dynamic Routing Algorithms*

In *static algorithms,* given fractions of the traffic (or given probabilities of routing) at node i of the network destined for each of the other nodes $j \neq i$, are directed on each of the outgoing links from node i. These fractions (probabilities) are decided upon before the network starts operation, are fixed in time and depend only on the time and ensemble averages of the message flow requirement.

In the case of *quasistatic routing,* the changes in these fractions (or probabilities) take place at given intervals of time and/or whenever extreme conditions (eg. link failure) occur. The time intervals between these changes will be long which ensures that messages are delivered in order and will not need individual addressing.

Dynamic routing allows continuous changes in routes depending upon the instantaneous system states and traffic conditions.

A further classification of *dynamic routing algorithms* is as follows:

- *Time dependent routing algorithms*
- *State dependent routing algorithms*

State dependent algorithms are also called *adaptive* algorithms.

Some authors [29] make a distinction between *dynamic routing* and *adaptive routing* in the following way. A *dynamic routing* is one wherein a part of the routing varies over time, while an *adaptive routing* is one where some part of the routing is a function of some estimate of the network state at the time a decision must be made. As per this classification, a *dynamic routing* is not necessarily *adaptive;* however the converse is unlikely. Implicit in the notion of *adaptive routing* is the notion of measurement of the network state, which is not required for purely *dynamic routing.*

Another frequently used classification based on the location in the network where the routing computation takes place is as follows:

- *Centralised Routing Algorithms*
- *Distributed Routing Algorithms*

In *centralised algorithms*, all route choices are made at a central node, while in *distributed algorithms* the computation of routes is shared among the network nodes with information exchanged between them as necessary. Note that this classification relates mostly to the implementation of the algorithm.

A routing algorithm can be viewed as the network layer protocol that guides packets through the communication subnet to their correct destination. Depending on how this protocol is implemented, the following classification can be made:

- *Datagram*
- *Virtual Circuit*

In a *datagram* network, two successive packets of the same user pair may travel along different routes, and a routing decision is necessary for each individual packet. In a *virtual circuit* network, a routing decision is made when each *virtual circuit* is set up. The routing algorithm is used to choose the communication path for the *virtual circuit*. All packets of the *virtual circuit* subsequently use this path up to the time that the *virtual circuit* is either terminated or rerouted for some reason.

2.3.2 Routing Strategies in Circuit-Switched Networks

Circuit-switched telecommunication has evolved from plain old telephony towards integrated digital networks. This evolution has roused increased customer demand combined with higher quality-of-service requirements. Network management strategies, specifically routing strategies, have undergone drastic changes due to this in the last decade. In this subsection, we present a survey of the evolution of routing strategies in the context of circuit-switched networks.

Automatic routing, in its simplest form, has existed since the introduction of the first step-by-step exchange. The digits dialed by the subscriber were transmitted

directly into outgoing trunk selections. Each exchange stripped off one or more digits and the connection was progressively established to the called party's line. However, this form of routing, referred to as *direct routing* in Figure 2.2, made inefficient use of transmission facilities and was found to be impractical for toll calls where a significant number of exchanges may need to be traversed.

With the introduction of common control crossbar switching machines in the 1940s, the possibility for an exchange to choose a route based on trunk group loading status also emerged. With this, the first forms of *alternate routing* were introduced. The simplest form of *alternate routing* that was suggested made use of tandem exchanges in metropolitan networks. Each exchange would have a number of direct trunks to each other exchanges and would also have overflow trunks to the tandem. The tandem would connect to every exchange serving subscribers. This arrangement enabled significant trunking efficiencies to be achieved.

The *alternate routing* concept based on the overflow technique, still had technological limitations that required the guarantee of integrity between numbering, signaling, and routing. A serious issue was the phenomenon of call looping, wherein a call returns to one of the switching centers along its routing path. The existence of multiple route choices in *alternate routing* gave rise to this possibility and thus the selection of compatible sets of alternatives was important. To overcome this difficulty, the concept of *hierarchical routing* was introduced in the 1950s, where the selection of alternative paths were subject to hierarchical rules. Due to the limited measurement capabilities in the switches and lack of computing facilities, routing (the hierarchical rules for the alternatives) was determined at the design stage and remained fixed under normal conditions until the next design stage. *Fixed hierarchical routing* was thus established in the 50s and is still used worldwide.

The introduction of *hierarchical routing* based on fixed alternatives led to significant gains in toll network efficiency. However, there were the following limitations to the efficiency gains achievable [15,43]:

- As traffic routing is closely related to the network design in fixed hierarchical

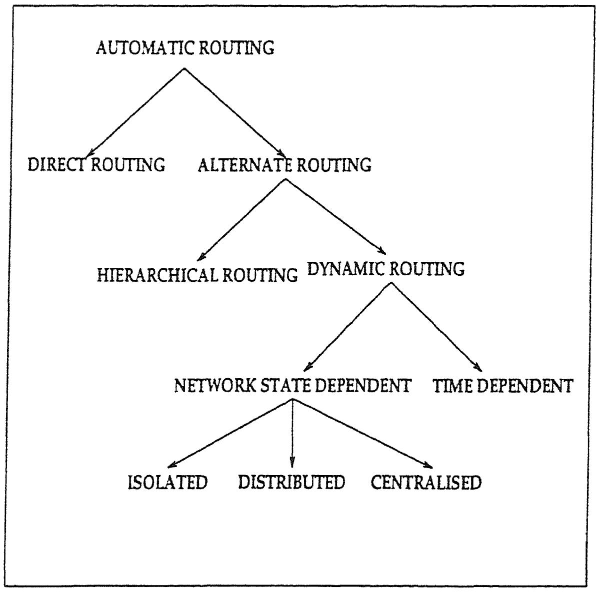

Figure 2.2: Relationship of Routing Methods in Circuit Switched Networks

networks, using a single routing pattern, determined according to busy-hour, busy-season traffic measurements, cannot allow the efficient accommodation of traffic in all situations. Analysis of the performance in fixed routed networks shows that there is some inconsistency between network capacity and the traffic demand, which leads to a poor grade of service in some parts of the network and unfairness between traffic streams. In many cases, there is an inefficient overprovision of the network.

- Network robustness is not ensured, also because of the hierarchical feature of routing. Little adaptivity in the overflow sequence makes it difficult to cope with situations other than nominal. Often, parts of the network with available capacity can not be used when other parts experience congestion. To address this problem, a range of network management controls were provided within the exchanges (e.g. re-routing within the hierarchy, set code blocks to stop traffic from entering the network for a particular destination, set trunk reservation levels in order to reserve a certain number of trunks for direct traffic

only). These were, in general, manually activated by the telephone company staff once they become aware of an overload situation. Inevitably, there will be time delays while these are being manually implemented, causing some loss of traffic.

New routing techniques referred to as *dynamic routing* were proposed in the 1980s in order to improve the network efficiency and to overcome the difficulties of the *hierarchical routing* mentioned above. The principle of *dynamic routing* is to profit from the existence of spare capacities in parts of the network while other parts are overloaded in order to maximise the use of network resources. The various methods proposed for *dynamic routing* may be classified against the following three parameters:

1. The basic mechanism used in the method:

   1a. *Time dependent schemes,* which take advantage of non-coincident busy periods across the network, on a pre-planned basis, but have limited ability to accommodate unplanned traffic fluctuations.

   1b. *Network state dependent schemes,* which sense the congestion patterns in the network and route traffic accordingly. They are also termed as *adaptive routing* strategies in the literature.

2. The topological scope of the routing calculations. The methods may be broadly grouped into three categories based on how large a portion of the network is taken into account in performing routing calculations:

   2a. *Centralised Routing algorithms* make network-wide routing calculations.

   2b. *Distributed Routing algorithms* perform calculations relating to a limited section of the network.

   2c. *Isolated Routing algorithms* consider only the state of the individual exchange (processor loading, trunk group status, etc.) and the patterns of recently received calls.

3. The frequency with which the routing recommendations are updated:

   3a. *Hours.*

   3b. *Minutes or less.*

   3c. *On a per call basis.*

The various routing strategies and their relationships is shown in Figure 2.2. Presently, *dynamic routing* is implemented in various networks such as the metropolitan [44], regional [45], long distance [46] and international [47] networks. Some of the major technological growths which will have impact upon *dynamic routing* can be stated as the following:

- The introduction of Fibre optic transmission systems as a transmission media may cause significant changes in the philosophy of network design. The almost unlimited capacity could be used to improve the network resiliency by providing a high level of redundancy. The implementation of the various functions of the *dynamic routing* like the transmission of the information regarding the network states to the network control center(s) (where the routing computation takes place) is expected to be significantly much faster.

- The current limitations placed upon *dynamic routing* methods lie primarily within the amount of data which must be processed to perform complex call routing on a network basis. This constraint will be reduced with the continuing improvements in computer technology.

Thus, to summarise the above discussion, we see that there is a shift in trend towards the implementation of *dynamic routing* algorithms in the context of circuit-switched networks also.

2.4 Flow Models in the Context of Routing

As mentioned in Chapter 1, the work on flow models was motivated by the need to have a modelling scheme that can conveniently be used for control synthesis.

From the discussion in the Subsection 2.2.6, we note that, a basic feature of the flow models is that they describe the time dependent average quantities of the network by means of deterministic ordinary differential equations. Note that the model assumes no stochastic properties of the input traffic and thus is equally applicable to both stationary and non-stationary traffic conditions. Based on the work of Filipiak in [21], we now examine how the model provides a control oriented network description.

Remark 2.5.1 In the discussion that follows we assume that packets are never rejected from any link.

Let $N = \{i, j, k, ..\}$ be the set of all nodes of a network, and let $L = \{(i, j), (j, k), ...\}$ denote the set of all unidirectional links (without loss of generality, a bidirectional link can be viewed as being constituted of two unidirectional links each possibly carrying traffic in opposite directions.) Let us define the set $O(j)$ as the set of all links of the network leaving node j and $I(j)$ as the set of all links entering node j. In other words,

$$\begin{aligned} O(j) &= \{(j, k) : (j, k) \in L\} \\ I(j) &= \{(i, j) : (i, j) \in L\}. \end{aligned}$$

Let $\lambda_j(t)$ denote the traffic entering the node j from outside the network. The traffic arriving at node j from all the links entering it, is given as $\sum_{i \in I(j)} \mu_{ij} G_{ij}(x_{ij})$. Therefore the total traffic arriving at node j is given as

$$\lambda_j^{total}(t) = \lambda_j(t) + \sum_{i \in I(j)} \mu_{ij} G_{ij}(x_{ij})$$

Let $\alpha_{jk}(t)$ denote the fraction of the total traffic arriving at node j that is routed onto link (j, k) at time t. Then the Equation (2.15) for the link (j, k) of the network can be written as

$$\begin{aligned} \dot{x}_{jk}(t) &= -\mu_{jk} G_{jk}(x_{jk}(t)) + \alpha_{jk}(t) \lambda_j^{total}(t) && (2.20) \\ &= -\mu_{jk} G_{jk}(x_{jk}) + \alpha_{jk}(t)(\lambda_j(t) + \sum_{i \in I(j)} \mu_{ij} G_{ij}(x_{ij})), \quad \forall (j, k) \in L. && (2.21) \end{aligned}$$

Let $\underline{X}(t)$ denote the state of the network, i.e.,

$$\underline{X} = \begin{bmatrix} \vdots \\ x_{jk}(t) \\ \vdots \end{bmatrix}$$

where $x_{jk}(t)$ $(\forall (j,k) \in L)$ denotes the average number of packets in the buffer associated with link (jk) at time t. Let $\underline{\alpha}(t)$ denote the vector of routing variables. i.e.

$$\underline{\alpha} = \begin{bmatrix} \vdots \\ \alpha_{jk}(t) \\ \vdots \end{bmatrix}$$

and let $\underline{\lambda}(t)$ denote the vector of input traffic to the network. i.e.

$$\underline{\lambda} = \begin{bmatrix} \lambda_1(t) \\ \lambda_2(t) \\ \vdots \\ \lambda_n(t) \end{bmatrix}$$

$\lambda_j(t)$ is the external traffic arriving at node j $(\forall j \in N)$. Then Equation (2.21) can be rearranged to get the following form:

$$\dot{\underline{X}}(t) = \underline{f}(\underline{X}, \underline{\alpha}, \underline{\lambda}(t)) \tag{2.22}$$

where the vector function $\underline{f}$ is given as

$$\underline{f} = \begin{bmatrix} \vdots \\ f_{jk} \\ \vdots \end{bmatrix}$$

for all $(j,k) \in L$. By comparing with Equation (2.21) it can be easily concluded that the scalar function $f_{jk}(\underline{X}, \underline{\alpha}, \underline{\lambda})$ is given as

$$f_{jk}(\underline{X}, \underline{\alpha}, \underline{\lambda}) \;=\; -\mu_{jk}G_{jk}(x_{jk}) + \alpha_{jk}(t)(\lambda_j(t) + \sum_{i \in I(j)} \mu_{ij}G_{ij}(x_{ij})). \tag{2.23}$$

The fraction of the traffic (at any node) that is routed onto an outgoing link is a real number in the interval $[0, 1]$. At any node in the network, since there is no accumulation of traffic, the sum of these fractions for all the outgoing links is unity. Therefore the routing variables $\alpha_{jk}(t)$ in the above equation must satisfy the following normalising conditions:

$$0 \leq \alpha_{jk}(t) \leq 1.$$
$$\sum_{k \in O(j)} \alpha_{jk}(t) = 1, \forall (j,k) \in L.$$

Observe that the description of the network given by Equation (2.22)

1) is applicable to networks of any topology and dimension.

2) makes no assumptions on the stochastic properties for the traffic vector $\underline{\lambda}(t)$. Therefore this description is applicable to both stationary and non-stationary traffic conditions.

3) allows the vector of routing variables $\underline{\alpha}(t)$ to be viewed as a **control vector**. Thus the problem of routing is amenable to control-theoretic formulation.

For any given input traffic, the choice of the vector of routing variables $\underline{\alpha}(t)$, determines the dynamical evolution of the state of the network (i.e. the mean buffer occupancies corresponding to each link). Consider an optimal routing strategy which minimises the total buffer occupancy time J defined as

$$J = \int_0^T \sum_{(j,k) \in L} x_{jk}(t) dt \qquad (2.24)$$

where T is the total duration of operation of the network. The problem of synthesising such an optimal routing strategy for the network can be stated mathematically as follows:

For the dynamical system given by Equation (2.22), with given initial conditions $\underline{X}(0) = \underline{X}_0$, and known input vector $\underline{\lambda}(t)$, find the control vector (of routing variables) $\underline{\alpha}(t)$ whose components satisfy the normalising conditions, such that the performance functional J given by the Equation (2.24) is minimised.

Thus we note that the possibility of describing the network as a dynamical system (with the routing variables as control variables) in the framework of flow models allows us to formulate the problem of optimal routing in a network as an optimal control problem. This formulation is applicable to networks of any topology and dimension and is equally applicable to both stationary and non-stationary traffic conditions. Using a familiar technique of optimal control theory, namely Pontryagin's Maximum Principle, the necessary conditions that the optimal solution to this problem has to satisfy can be obtained. In the chapters that follow, we apply this technique to the two node network, three node network and larger networks that we mentioned in Chapter 1.

Chapter 3

Optimal and Suboptimal Routing Strategies for a Two Node Network

3.1 Introduction

In the previous chapter, we reviewed the framework of dynamic flow models within which the problem of optimal routing was formulated as an optimal control problem. Our emphasis in Chapter 2 was on the work done by Filipiak [21]. In [21] the output flow intensity from each link of a network is assumed to be governed by an empirically determined[1] non-linear function of the average number of packets/customers in the buffer associated with it. In a packet switching context, as the flow on a link can not exceed the channel capacity, it is reasonable to assume that this function saturates at the capacity value. It is also reasonable as in [21], to assume that the flow out from a buffer will be zero when it is empty and will increase with increasing buffer occupancy. Based on these considerations, we assume in this chapter and in the subsequent chapters, that the flow out function depends linearly on the average number of packets in the buffer and has an upper bound equal to the channel capacity of the link. We also assume that the buffers

[1] In some specific cases such as an M/M/1 queue, it is possible to obtain this function from queueing theoretic considerations. However in majority of the cases, this function has to be determined from measurements.

are of infinite capacity and therefore packets are never rejected from any link.

The network topology which we consider in this chapter is a simple one of two nodes as shown in Figure 3.1. Packets arriving at node 1 are to be routed to the destination node 2 through one of the two available paths, namely link 1 and link 2. In the analysis that follows, we assume that link 1 has a finite channel capacity equal to C_1 and link 2 has infinite channel capacity.

Let $x_1(t)$ and $x_2(t)$ denote the number of packets at time t, in the buffers associated with links 1 and 2 respectively. Let $\alpha_1(t)$ be the fraction of traffic (of the total traffic $\lambda(t)$ arriving at node 1) routed on link 1 at time t, and $\alpha_2(t)$ be that on link 2. Then, from flow conservation principle and from the assumption regarding the nature of the flow out functions, it follows that the dynamic evolution of the state of the network is given by the following differential equations:

$$\dot{x}_1(t) = -f_1(x_1) + \alpha_1(t)\lambda(t) \tag{3.1}$$

$$\dot{x}_2(t) = -a_2 x_2 + \alpha_2(t)\lambda(t) \tag{3.2}$$

where the function $f_1(x_1)$ is specified as follows:

$$f_1(x_1) = \begin{cases} a_1 x_1 & \text{if } x_1 < x_{1s} = C_1/a_1 \\ C_1 & \text{otherwise} \end{cases} \tag{3.3}$$

3.2 Problem Formulation

The problem of synthesizing the optimal routing strategy is formulated as follows:

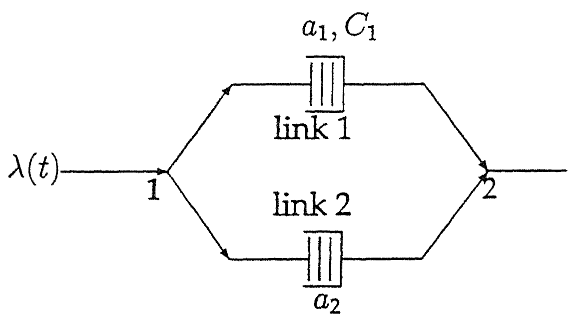

Figure 3.1: Network Topology of two nodes

Given the network parameters a_1, a_2 and C_1, and the traffic pattern $\lambda(t)$ during the interval of operation $[0, T]$ for the system whose dynamics is given by (3.1) and (3.2), find the routing variables $\alpha_1(t)$ and $\alpha_2(t)$ ($\forall t \in [0, T]$), which satisfy the normalizing conditions:

$$0 \leq \alpha_1(t) \leq 1, \tag{3.4}$$

$$0 \leq \alpha_2(t) \leq 1, \tag{3.5}$$

$$\alpha_1(t) + \alpha_2(t) = 1, \quad \forall t \in [0, T] \tag{3.6}$$

so as to minimise the performance index

$$J = \int_0^T (x_1(t) + x_2(t))dt \tag{3.7}$$

The above performance index J is the total buffer occupancy time and the rationale for minimizing this is that a waiting cost is incurred at a rate proportional to the number of customers present in the system.

3.3 Solution Approach

Assumption 3.3.1 In the analysis that follows, we assume that the initial buffer occupancies of the network do not exceed the saturation values, i.e. $x_1(0) \leq x_{1s}$ and $x_2(0)$ is finite.

Remark 3.3.1 The assumption that the initial buffer occupancy $x_1(0)$ does not exceed the saturation value x_{1s} may appear to be overly restrictive, but is not quite so in practice as can be argued on the following grounds. Consider the network to be operated for finite intervals $[0, T_1]$, $[T_2, T_3]$, $[T_4, T_5]$ ...., (the network is assumed to be shut down during (T_1, T_2), (T_3, T_4),....) which is the situation considered for investigation in this thesis. We shall soon prove that, the optimal routing strategy or the suboptimal routing strategy which we propose in Section 3.5 of this chapter has the following interesting property. If at the beginning of an interval of network operation say $[T_i, T_j]$, $x_1(T_i) \leq x_{1s}$, then under the optimal/suboptimal routing

strategy, $x_1(T_j) \leq x_{1s}$. Since it is reasonable to assume that at the very beginning of the network operation, the buffer occupancies are zero, and furthermore during the intervals of shutdown, the buffer occupancies can not increase (since no traffic is admitted into the network), the assumption that at the beginning of any interval of operation, the buffer occupancy of link 1 does not exceed the saturation value x_{1s}, is not restrictive in practical terms.

To solve the above optimal control problem, we use the maximum principle by Pontryagin [22,23]. The function $\frac{df_1(x_1)}{dx_1}$ is discontinuous at $x_1 = x_{1s}$. For the maximum principle to be applicable it is necessary that $\frac{df_1(x_1)}{dx_1}$ be continuous in x_1. We therefore consider a class of optimal control problems in which the dynamics of link 1 is modified as

$$\dot{x}_1 = -f_1^r(x_1) + \alpha_1(t)\lambda(t) \tag{3.8}$$

The function $f_1^r(x_1)$ (shown in Figure 3.2) is defined as below:

$$f_1^r(x_1) = \begin{cases} a_1 x_1 & \text{if } x_1 \leq x_l = \frac{C_1}{a_1} + \frac{r(1-\sqrt{1+a_1^2})}{a_1(\sqrt{1+a_1^2})} \\ \sqrt{r^2 - (x_1 - x_s)^2} + C_1 - r & \text{if } x_l < x_1 < x_s \\ C_1 & \text{if } x_1 \geq x_s \end{cases} \tag{3.9}$$

where, $x_s = \frac{C_1}{a_1} + \frac{r(\sqrt{(1+a_1^2)}-1)}{a_1}$ and $r > 0$.

$$\frac{df_1^r(x_1)}{dx_1} = \begin{cases} a_1 & \text{if } x_1 \leq x_l \\ \frac{(x_s - x_1)}{\sqrt{r^2-(x_s-x_1)^2}} & \text{if } x_l \leq x_1 \leq x_s \\ 0 & \text{if } x_1 > x_s \end{cases} \tag{3.10}$$

Notes:

1. The function $f_1^r(x_1)$ is obtained by drawing an arc of radius r, tangential to the lines $f_1(x_1) = a_1 x_1$ and $f_1(x_1) = C_1$, as shown in Figure 3.2. The centre of this arc is at $(x_s, C_1 - r)$.

2. As $r \to 0$, $f_1^r(x_1) \to f_1(x_1), \forall x_1 \geq 0$; $x_l \to x_{1s}$; $x_s \to x_{1s}$.

3. From Equation (3.10), it can easily be verified that $\frac{df_1^r}{dx_1^r}$ is a monotonically decreasing function in the interval $[x_l, x_s]$.

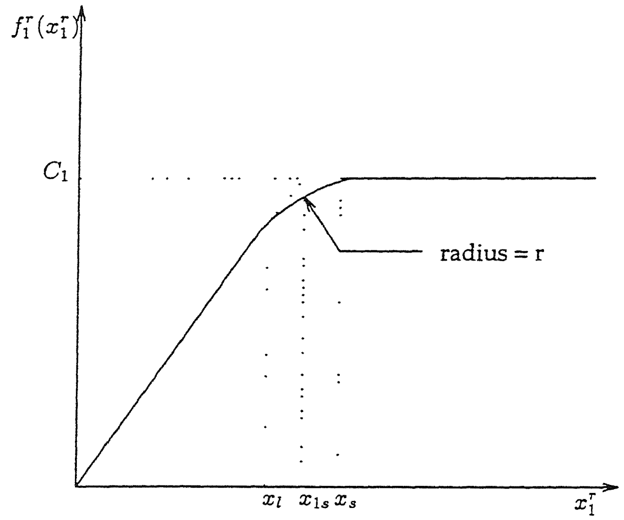

Figure 3.2: Modified Flow out function

4. The dynamics of link 2, the normalizing conditions on $\alpha_1(t)$ and $\alpha_2(t)$ and the performance index J are the same as that of the original formulation. The earlier Assumption **3.3.1** that the initial buffer occupancy $x_1(0) \leq x_{1s}$ is modified as, $x_1^r(0) \leq x_l$.

For the above class of problems, we investigate the nature of the optimal routing strategy. Let $x_1^r(t)$ and $x_2^r(t)$ be the optimal state variables, $p_1^r(t)$ and $p_2^r(t)$ be the optimal costate variables and $\alpha_1^r(t)$ and $\alpha_2^r(t)$ be the optimal control variables corresponding to a choice r. The Hamiltonian $H^r(\underline{x}^r, \underline{p}^r, \underline{\alpha}^r)$ is given as:
$H^r(\underline{x}^r, \underline{p}^r, \underline{\alpha}^r) = x_1^r(t) + x_2^r(t) - p_1^r f_1^r(x_1^r) + p_1^r \alpha_1^r \lambda(t) - p_2^r a_2 x_2^r + p_2^r \alpha_2^r \lambda(t)$.
Minimizing H^r w.r.t. $\alpha_1^r(t)$ and $\alpha_2^r(t)$ yields the following optimal control policy:

- If $p_1^r(t)$ is less than $p_2^r(t)$, then $\alpha_1^r(t)$ equals unity and $\alpha_2^r(t)$ equals zero.
- If $p_1^r(t)$ is greater than $p_2^r(t)$, then $\alpha_1^r(t)$ equals zero and $\alpha_2^r(t)$ equals unity.
- If for all t in some interval I, $p_1^r(t)$ and $p_2^r(t)$ are identically the same, then $\alpha_1^r(t)$ is not specified by the above conditions.

According to the maximum principle, the optimal costate variables $p_1^r(t)$ and $p_2^r(t)$

must satisfy the following differential equations:

$$\begin{aligned}\dot{p}_1^r &= -\partial H^r/\partial x_1^r && (3.11)\\ &= -1 + \frac{df_1^r}{dx_1^r} p_1^r(t) \\ \dot{p}_2^r &= -\partial H^r/\partial x_2^r && (3.12)\\ &= -1 + a_2 p_2^r(t)\end{aligned}$$

along with the transversality conditions $p_1^r(T) = p_2^r(T) = 0$.
The solution for $p_2^r(t)$ is given as

$$p_2^r(t) = \frac{1 - e^{a_2(t-T)}}{a_2} \tag{3.13}$$

Remark 3.3.2 If a_1 is less than a_2, then it can be argued that $\alpha_1^r(t)$ equals zero and $\alpha_2^r(t)$ equal unity, $\forall t \in [0, T]$ is the optimal routing strategy. This follows from the fact that

$$p_1^r(t) = \frac{(1-e^{a_1(t-T)})}{a_1} > p_2^r(t) = \frac{(1-e^{a_2(t-T)})}{a_2}, \forall t \in [0, T).$$

Remark 3.3.3 If a_1 equals a_2, then it can be argued that any routing strategy which ensures that $x_1^r(t)$ never exceeds x_l [in the limiting case as r tends to zero, x_{1s}] achieves optimality.

In what follows, we therefore focus our attention to the case wherein $a_1 > a_2$. Let x_{tr} be the value of x_1^r at which $\frac{df_1^r}{dx_1^r}$ equals a_2. Then $x_l < x_{tr} < x_s$ and as r tends to zero, x_{tr} tends to x_{1s}.

Before proceeding to investigate the nature of the optimal routing strategy for this class of models, and discuss the implementation details, we first define below some of the terms which are frequently used in the remaining part of this chapter.

Definition 3.3.1 If for all t in some interval I, $x_1^r(t) \leq x_l$ then the network is said to be operating in the linear mode during I, and such an interval I is termed as a linear regime.

Definition 3.3.2 If for all t in $I, x_1^r(t) \geq x_s$ then the network is said to be operating in the saturation mode during I and such an interval I is termed as a saturation regime.

Definition 3.3.3 If for all t in I, $x_l < x_1^r(t) < x_s$ then the network is said to be operating in the transition mode during I and such an interval I is termed as a transition regime.

For the model corresponding to the limiting case in which r tends to zero, the above terms are defined as follows:

Definition 3.3.4 If for all t in $I, x_1(t) < x_{1s}$, then I is termed as a linear regime and the network is said to be operating in the linear mode during I.

Definition 3.3.5 If for all t in $I, x_1(t) = x_{1s}$, then I is termed as a transition regime and the network is said to be operating in the transition mode during I.

Definition 3.3.6 If for all t in $I, x_1(t) > x_{1s}$, then I is termed as a saturation regime and the network is said to be operating in the saturation mode during I.

Lemma 3.3.1 *If* $\forall t \in I, p_1^r(t) = p_2^r(t)$, *then* $\alpha_1^r(t) = \frac{f_1^r(x_{tr})}{\lambda(t)}$ *during* I.

Proof :

$$\begin{aligned}
& p_1^r(t) && \equiv && p_2^r(t) \\
\Rightarrow\ & \dot{p}_1^r(t) && \equiv && \dot{p}_2^r(t) \\
\Rightarrow\ & \frac{df_1^r(x_1^r)}{dx_1^r} && \equiv && a_2 \\
\Rightarrow\ & x_1^r(t) && \equiv && x_{tr} \text{ and} \\
& \dot{x}_1^r(t) && \equiv && 0 \ \forall t \in I.
\end{aligned}$$

Therefore $\alpha_1^r(t) = \frac{f_1^r(x_{tr})}{\lambda(t)}, \forall t \in I$. In the limiting case as r tends to zero, $\alpha_1^r(t) = \frac{C_1}{\lambda(t)}$.

□

Definition 3.3.7 An interval during which $\alpha_1^r(t)$ equals $\frac{f_1^r(x_{tr})}{\lambda(t)}$ is termed as an *interval of partial routing*.

Lemma 3.3.2 *The functions* $p_1^r(t)$ *and* $p_2^r(t)$ *are non-negative in the interval* $[0, T]$.

Proof : It is easy to verify that the function $p_2^r(t)$ which is equal to $\frac{(1-e^{a_2(t-T)})}{a_2}$ is non-negative over $[0, T]$.

$$\dot{p}_1^r = -1 + \frac{df_1^r}{dx_1^r} p_1^r, \quad p_1^r(T) = 0.$$

Solving the above,

$$p_1^r(t) = [p_1^r(0) - \int_0^t e^{-\int_0^\tau \frac{df_1^r(x_1^r(\gamma))}{dx_1^r} d\gamma} d\tau][e^{\int_0^t \frac{df_1^r(x_1^r(\tau))}{dx_1^r} d\tau}]$$

$$\text{where} \quad p_1^r(0) = \int_0^T e^{-\int_0^\tau \frac{df_1^r(x_1^r(\gamma))}{dx_1^r} d\gamma} d\tau > \int_0^t e^{-\int_0^\tau \frac{df_1^r(x_1^r(\gamma))}{dx_1^r} d\gamma} d\tau \quad \text{for } 0 \le t < T.$$

Since $p_1^r(t)$ is the product of two positive terms, it is positive $\forall t \in [0, T)$. Hence the lemma.

□

Towards finding the nature of the optimal routing strategy, we prove the following theorems.

Theorem 3.3.1 *The system can not end in :*

(a) a saturation regime

(b) a transition regime wherein $\forall t, x_1^r(t) > x_{tr}$.

Proof : In either of the above two cases, $x_1^r(t) > x_{tr}$. Let $I_1 = [t_1, T]$ be an interval during which $x_1^r(t) > x_{tr}$. Then $\frac{df_1^r}{dx_1^r} < a_2$ over the entire interval I_1. Therefore, $\dot{p}_1^r < -1 + a_2 p_1^r$ (since by Lemma 3.3.2, $p_1^r(t) \ge 0, \forall t \in [0, T]$) while $\dot{p}_2^r = -1 + a_2 p_2^r$ in I_1.

It can then be argued[2] that, $p_1^r(t) > p_2^r(t)$ in $[t_1, T)$. Consequently $\alpha_1^r(t) \equiv 0$ in I_1. The dynamics of $x_1^r(t)$ in I_1 is given as:

$$\dot{x}_1^r = -f_1^r(x_1^r).$$

Therefore $x_1^r(t)$ is a monotonically decreasing function of time in I_1, and at $t = t_1$,

$$\begin{aligned} x_1^r(t_1) &= \int_{t_1}^T f_1^r(x_1^r(\tau)) d\tau + x_1^r(T) \\ &> (T - t_1) f_1^r(x_{tr}) + x_{tr} \quad \text{(as per the assumption that } x_1^r(\tau) > x_{tr}, \forall \tau \in I_1.) \end{aligned}$$

[2]We do not go into these arguments for reasons of brevity.

From the above property of $x_1^r(t)$ (that this function is monotonically decreasing in I_1 and everywhere greater than x_{tr}) and the from the fact[3] that $x_1^r(t)$ is a continuous function in the entire interval $[0, T]$, it follows that I_1 is preceded by an interval in which $x_1^r(t) > x_{tr}$. Let us denote this interval by I_0 where $I_0 = [t_0, t_1]$ (for some $t_0 < t_1$). Then the arguments which were used to prove that $x_1^r(t)$ is monotonic and decreasing in I_1 can now be extended to the interval $I_0 \bigcup I_1 = [t_0, T]$ and it can be argued that I_0, in turn, is preceded by an interval during which $x_1^r(t) > x_{tr}$. Extending these arguments, it can be concluded that $x_1^r(t) > x_{tr}, \forall t \in [0, T]$, and in particular $x_1^r(0) > x_{tr} + \int_0^T f_1^r(x_1^r(\tau))d\tau$. Since this violates the assumption that the initial buffer occupancy $x_1^r(0) \leq x_l$, the terminating regime can not be one in which, for all t, $x_1^r(t) > x_{tr}$.

□

In the case of the model in which r tends to zero, if the network operation starts in the linear mode, then the buffer occupancy of link 1 (i.e. $x_1(t)$) never exceeds a value equal to $x_{1_s} + C_1(T - t)$. This value is thus an upper bound (not necessarily tight) on $x_1(t)$ provided the routing strategy is optimal.

Theorem 3.3.2 *The value of the routing variable $\alpha_1^r(t)$ is either* 1 *or* $\frac{f_1^r(x_{tr})}{\lambda(t)}$, $\forall t \in [0, T]$.

Proof : We show that $p_1^r(t) \leq p_2^r(t), \forall t \in [0, T]$, as follows:
Let us assume on the contrary that $p_1^r(t) > p_2^r(t) = \frac{1-e^{a_2(t-T)}}{a_2}$, for all t in some interval $I_0 = [t_0, t_1)$. Then $\alpha_1^r(t) \equiv 0$ over I_0, and $x_1^r(t)$ is a monotonically decreasing function in I_0.
By reasonings similar to those used in the proof of Theorem 3.3.1, it can be argued that if there is any subinterval in I_0 during which $x_1^r(t) > x_{tr}$, then it should have been preceded by such an interval (wherein $x_1^r(t) > x_{tr}$) all the way starting at $t = 0$. Since this results in the violation of the initial condition $x_1^r(0) \leq x_l$, it can be concluded that there exists no such subinterval of I_0 during which $x_1^r(t) > x_{tr}$. Consequently $x_1^r(t) \leq x_{tr}$, $\forall$ t $\in I_0$.

[3]Implicit in this inference is the assumption that the load pattern doesn't have any impulses and therefore the buffer occupancies are continuous functions of time.

Hence $\frac{df_1^r}{dx_1^r} \geq a_2 \; \forall t \in I_0$. During I_0,

$$\begin{aligned} \dot{p_1^r} &\geq -1 + a_2 p_1^r \quad \text{and} \\ \dot{p_2^r} &= -1 + a_2 p_2^r \\ (\dot{p_1^r} - \dot{p_2^r}) &\geq a_2(p_1^r - p_2^r) \\ &> 0 \quad \text{(as per the assumption that } p_1^r(t) > p_2^r(t), \forall t \in I_0.) \end{aligned}$$

Thus $(p_1^r(t) - p_2^r(t))$ is not only positive, but also monotonically increasing in I_0, and hence[4] $p_1^r(t_1) - p_2^r(t_1) > p_1^r(t_0) - p_2^r(t_0) > 0$. This implies that I_0 is followed by an interval $[t_1, t_2)$ in which $p_1^r(t) > p_2^r(t)$. The arguments that were used to prove that $p_1^r(t) - p_2^r(t)$ is monotonically increasing in I_0 can now be extended to $I_0 \bigcup I_1$ and it can be argued that $I_0 \bigcup I_1$ is also followed by an interval in which $p_1^r(t) > p_2^r(t)$. Extending these arguments, it can be concluded that the relationship $p_1^r(t) > p_2^r(t)$ has to hold true for all t in $[t_0, T]$. But $p_1^r(T) > p_2^r(T)$ violates the transversality condition and therefore the initial assumption of the existence of an interval I_0 during which $p_1^r(t) > p_2^r(t)$ is incorrect. We have already seen that if $p_1^r(t) < p_2^r(t)$, then $\alpha_1^r(t) = 1$. And if for all t in some interval I, $p_1^r(t) = p_2^r(t)$ then $\alpha_1^r(t) = \frac{f_1^r(x_{tr})}{\lambda(t)}$, in I (by Lemma 3.3.1). Thus the routing variable $\alpha_1^r(t)$ is either 1 or $\frac{f_1^r(x_{tr})}{\lambda(t)}$ for all $t \in [0, T]$.

□

From the Theorems 3.3.1 and 3.3.2, an interesting observation on the nature of the optimal routing strategy emerges. If during the network operation, the buffer occupancy of link 1 reaches a value equal to x_{tr} (in the limiting case as r tends to zero, the value equal to x_{1s}) at an instant of time t_0, and if the input traffic $\lambda(t)$ to the network for the remaining interval of network operation $[t_0, T]$ is greater than $f_1^r(x_{tr})$ (C_1 in the limiting case), then the optimal routing strategy for that interval $[t_0, T]$ is given as $\alpha_1^r(t) = \frac{f_1^r(t)}{\lambda(t)}$ (in the limiting case $\frac{C_1}{\lambda(t)}$). In other words, the remaining time span of network operation is an *interval of partial routing* in which both the links 1 and 2 are used to carry the arriving traffic $\lambda(t)$.

[4]Implicit in this inference is the fact that the costate variables are continuous functions of time.

During an *interval of partial routing*, $\alpha_1^r(t)$ equals $\frac{f_1^r(x_{tr})}{\lambda(t)}$ and $x_1^r(t)$ is identically equal to x_{tr}. We now show that such an interval, can neither be followed by an interval in which $x_1^r(t)$ is less than x_{tr}, nor be preceded by one in which $x_1^r(t)$ is greater than x_{tr}.

Theorem 3.3.3 *An interval of partial routing can not be:*

(a) followed by a transition regime in which $x_1^r(t)$ is less than x_{tr}

(b) preceded by a transition regime in which $x_1^r(t)$ is greater than x_{tr}.

Proof : Let $I = [t_0, t_1]$ be an *interval of partial routing*. During $I, p_1^r(t)$ is identically equal to $p_2^r(t)$.

(a) Assume that I is followed by an interval I_1 during which $x_1^r(t) < x_{tr}$. Then $\dot{p}_1^r > -1 + a_2 p_1^r$ and $\dot{p}_2^r = -1 + a_2 p_2^r, \forall t \in I_1$.

$$\begin{aligned} \text{This implies that} \quad (\dot{p}_1^r - \dot{p}_2^r) &> a_2(p_1^r - p_2^r), \quad \forall t \in I_1. \\ \text{Therefore} \quad (p_1^r(t) - p_2^r(t)) &> (p_1^r(t_1) - p_2^r(t_1))e^{a_2(t-t_1)}, \quad \forall t \in I_1. \\ \text{Hence} \quad (p_1^r(t) - p_2^r(t)) &> 0, \quad \forall t \in I_1. \end{aligned}$$

But it has already been proved that $p_1^r(t)$ is upper bounded by $p_2^r(t)$ (in the proof of Theorem 3.3.2) for all t in $[0, T]$. Thus the assumption that I is followed by an interval during which $x_1^r(t) < x_{tr}$ is incorrect.

(b) If I is preceded by an interval I_0 during which $x_1^r(t) > x_{tr}$, then $\dot{p}_1^r < -1 + a_2 p_1^r$ and $\dot{p}_2^r = -1 + a_2 p_2^r, \forall t \in I_0$, which implies that $(\dot{p}_1^r - \dot{p}_2^r) < a_2(p_1^r - p_2^r)$ in I_0. Since $p_1^r(t_0) - p_2^r(t_0) = 0$, it can be argued that $p_1^r(t) > p_2^r(t), \forall t \in I_0$, which is a contradiction as per Theorem 3.3.2. Thus an *interval of partial routing* can neither be followed by a transition regime in which $x_1^r(t)$ is less than x_{tr} nor be preceded by a regime in which $x_1^r(t)$ is greater than x_{tr}.

□

Comments: It can be seen that the above theorems and lemmas hold true even if the assumption on the initial buffer occupancy ($x_1^r(0) \leq x_l$) is relaxed to $x_1^r(0) \leq x_{tr}$. In the limiting case of r tending to zero, both x_l and x_{tr} tend to x_{1s}, and therefore, for the original model this relaxation is of no significance.

3.4 Optimal Routing Strategy: Implementation

Theorem 3.3.2 states that either the entire input traffic is routed onto the *faster* link, or there are *intervals of partial routing*. Therefore the optimal routing strategy can be implemented in the following steps:
I. Check if $\alpha_1^r(t) = 1, \forall t \in [0, T]$ is the optimal routing strategy by the following steps:

1. Integrate the equation $\dot{x}_1^r = -f_1^r(x_1^r) + \lambda(t)$ with the given initial condition $x_1^r(0)$ and obtain $x_1^r(t)$ and $\frac{df_1^r(x_1^r)}{dx_1^r}, \forall t \in [0, T]$.

2. Integrate the equation $\dot{p}_1^r = -1 + \frac{df_1^r}{dx_1^r}.p_1^r$ with $p_1^r(T) = 0$ from $t = T$ to $t = 0$, and obtain $p_1^r(t)$.

3. Check if $p_1^r(t) < \frac{1-e^{a_2(t-T)}}{a_2}, \forall t \in [0, T]$. If this inequality is satisfied for all t, then $\alpha_1^r(t) \equiv 1$ is the optimal routing strategy.

II. If the inequality in step I.3 is violated for some $t \in [0, T]$, then there are *intervals of partial routing* during which $\alpha_1^r(t)$ equals $\frac{f_1^r(x_{tr})}{\lambda(t)}$. (For the model in which r tends to zero, $\alpha_1(t)$ equals $\frac{C_1}{\lambda(t)}$).

We now investigate a procedure to specify the duration of such *intervals of partial routing*, if they exist. We shall first investigate the case wherein the traffic pattern $\lambda(t)$ has a single positive crossing over the value $f_1^r(x_{tr})$ (in the limiting case C_1) as shown in Figure 3.3 and Figure 3.4.
In the case of a load pattern $\lambda(t)$ which ends with a value greater than $f_1^r(x_{tr})$ (in the limiting case C_1) it has already been argued that the optimal routing strategy is specified as:

$$\alpha_1^r(t) = \begin{cases} 1 & ; \ \forall t \in [0, t_s] \\ \frac{f_1^r(x_{tr})}{\lambda(t)} & ; \ \forall t \in (t_s, T]. \end{cases}$$

where t_s is the first instant at which $x_1(t) = x_{tr}$.

For load patterns which end with a value less than $f_1^r(x_{tr})$ (in the limiting case C_1) as shown in Figure 3.4, it can be argued from Theorems 3.3.2 and 3.3.3 that, if

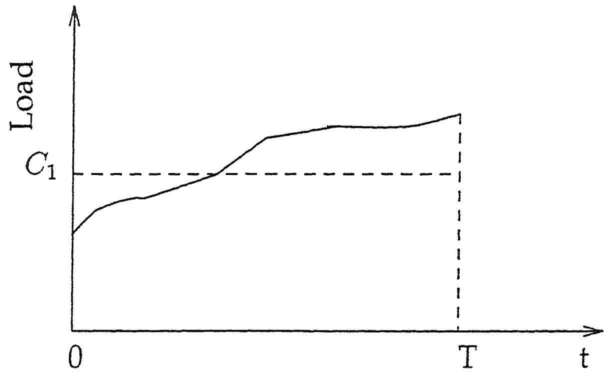

Figure 3.3: Load $\lambda_1(t)$ which ends with a value greater than C_1

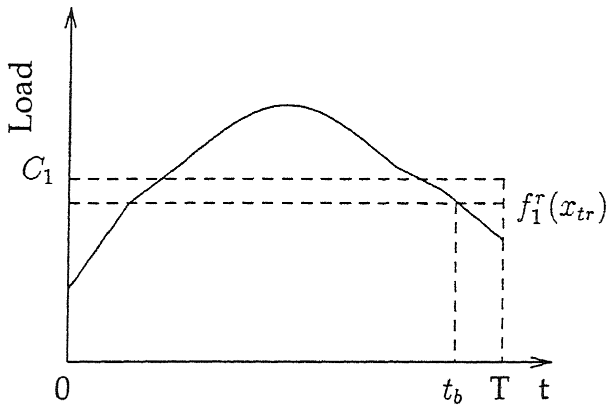

Figure 3.4: Load $\lambda_1(t)$ with a single positive crossing over the value C_1

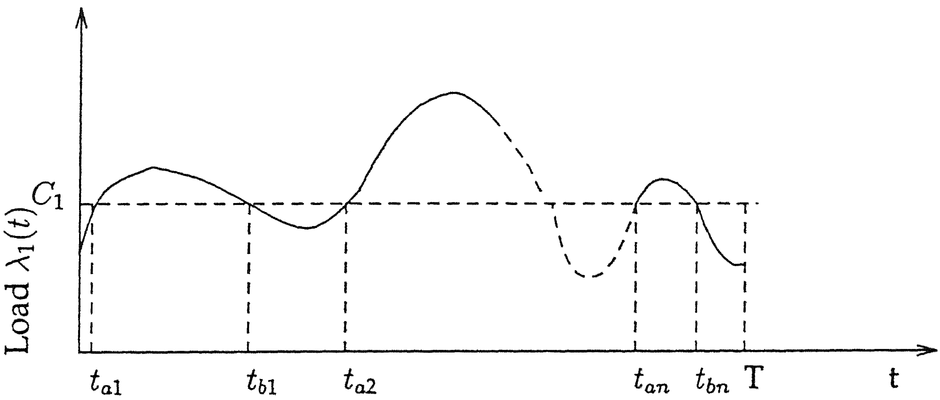

Figure 3.5: Load $\lambda_1(t)$ which has n positive crossings over the value C_1

there is an *interval of partial routing* then it has to start at the instant t_s (t_s is the first instant at which $x_1^r(t)$ equals x_{tr}) For $t \in [0, t_s]$, $\alpha_1^r(t) = 1$. Let the *interval of partial routing* be $[t_s, t_1]$ where $t_s < t_1 < t_b$. The optimal routing strategy is then given by:

$$\alpha_1^r(t) = \begin{cases} 1 & ; \ \forall t \in [0, t_s]. \\ \frac{f_1^r(x_{tr})}{\lambda(t)} & ; \ \forall t \in (t_s, t_1]. \\ 1 & ; \ \forall t \in (t_1, T]. \end{cases}$$

Solution for t_1:

$$p_1^r(t_1) = p_2^r(t_1) = \frac{1 - e^{a_2(t_1 - T)}}{a_2} \tag{3 14}$$

At $t = t_1$, $x_1^r(t_1)$ equals x_{tr}. Let t_1^* ($t_1^* > t_1$) be the instant when $x_1^r(t)$ reaches the value x_{tr}. During the interval $[t_1, t_1^*]$, $x_1^r(t) > x_{tr}$ and during $[t_1^*, T]$, $x_1^r(t) < x_{tr}$.

$$\int_{t_1}^{t_1^*} (\lambda(\tau) - f_1^r(x_1^r(\tau)))d\tau = x_1^r(t_1^*) - x_1^r(t_1) = 0 \tag{3 15}$$

In the control problem corresponding to the case where r tends to zero, Equation (3 15) reduces to the following:

$$\int_{t_1}^{t_1^*} \lambda(\tau)d\tau = C_1(t_1^* - t_1) \tag{3 16}$$

Furthermore, for the limiting case of r tending to zero, the dynamics of $p_1^r(t)$ over $[t_1, t_1^*]$ is given by

$$\begin{aligned} p_1^r(t) &= p_1^r(t_1) - (t - t_1) \\ &= \frac{1 - e^{a_2(t_1 - T)}}{a_2} - (t - t_1) \ \text{(after substituting the r.h s of Equation 3.14).} \end{aligned}$$

At $t = t_1^*$,

$$p_1^r(t_1^*) = \frac{1 - e^{a_1(t_1^* - T)}}{a_1} = \frac{1 - e^{a_2(t_1 - T)}}{a_2} - (t_1^* - t_1) \tag{3 17}$$

Solution to the Equations (3.16) and (3.17) gives the instants t_1 and t_1^*.

3.4.1 Numerical Examples

For a network with link parameters $a_1 = 0\ 9$, $a_2 = 0\ 2$, and $C_1 = 9$ and initial buffer occupancies $x_1(0) = x_2(0) = 0$, operated for an interval $[0, 10]$, the optimal routing strategies and the corresponding performance indices are obtained for various choices of the traffic $\lambda(t)$ by the following procedure:
Consider the performance index J of a routing strategy given below

$$\alpha_1(t) = \begin{cases} 1 & ; \ \forall t \in [0, t_s] \\ \frac{C_1}{\lambda(t)} & , \ \forall t \in (t_s, t_p] \\ 1 & ; \ \forall t \in (t_p, T] \end{cases}$$

The variation in the performance index J as the parameter t_p is varied over $[t_s, t_b]$ is plotted in each of the examples (shown in Figures 3.6, 3 7, 3 8 and 3.9) The value of t_p at which the performance index attains the minimum is the instant t_1 corresponding to the optimal routing strategy
It can be seen that for the load pattern in Example 1 (Figure 3 6), the minimum performance index is obtained at $t_p = t_s = 0\ 67$.
It is easily verified in each of the examples that the value of t_p at which J is minimum, satisfies the Equations (3 16) and (3.17) derived in the previous section.

3.4.2 Extension to Arbitrary Load Patterns

Consider a load pattern $\lambda(t)$ which has n positive crossings above the value equal to C_1 as shown in Figure 3 5. From the Theorems 3.3.2 and 3 3.3, it can be concluded that *intervals of partial routing*, if they exist, are of the type $[t_{s_1}, t_1], [t_{s_2}, t_2], \ldots [t_{s_n}, t_n]$ where $t_{a_i} \leq t_{s_i} \leq t_i < t_{b_i}$ and t_{s_1}, t_{s_2} are the instants at which $x_1(t) = x_{1s}$.

Notes:

1. In the arguments that follow, we consider the case wherein r tends to zero. Hence x_l, x_s and x_{tr} tend to $x_{1s} = \frac{C_1}{a_1}$.

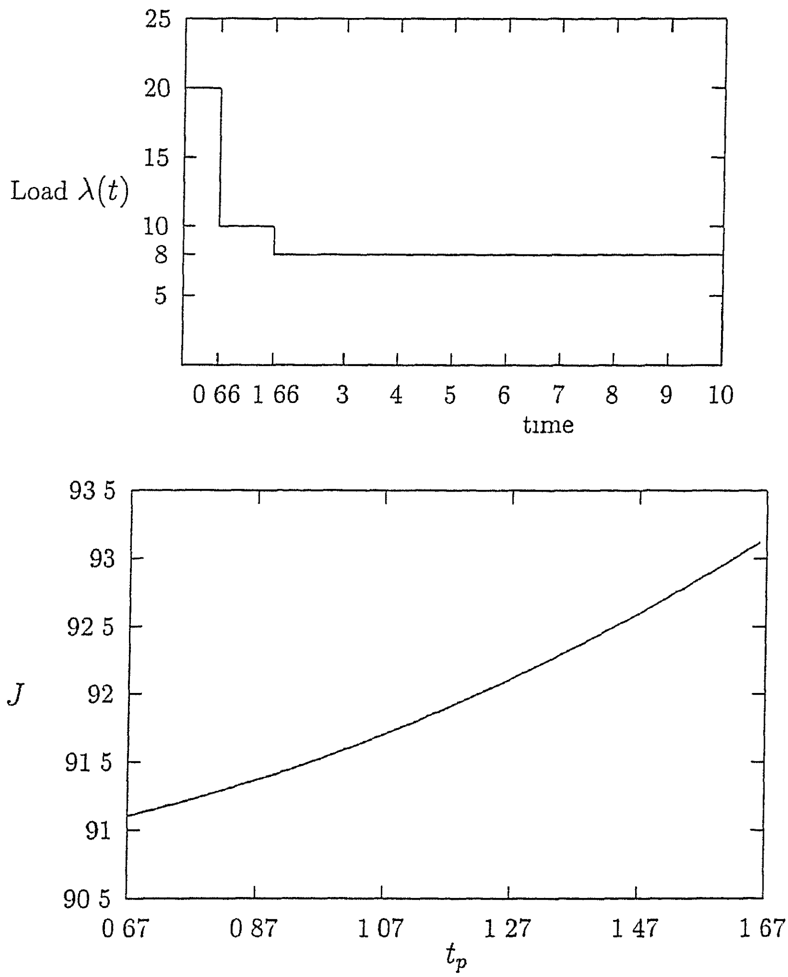

Figure 3 6: Performance vs. t_p for Example 1

The value of t_p at which the performance index is minimum = 0.67.

2. The instants $t_{s_1}, t_{s_2}, \ldots t_{s_n}$ are those for which the value of $x_1(t)$ reaches x_{1s} from a smaller value i e. $x_1(t_{s_i}^-) < x_{1s}$ and $x_1(t_{s_i}) = x_{1s}$.

The optimal routing strategy is specified as:

$$\alpha_1(t) = \begin{cases} \frac{C_1}{\lambda(t)} & \text{for the interval } [t_{s_i}, t_i],\ i = 1, 2, 3, \ldots\text{n.} \\ 1 & \text{elsewhere.} \end{cases}$$

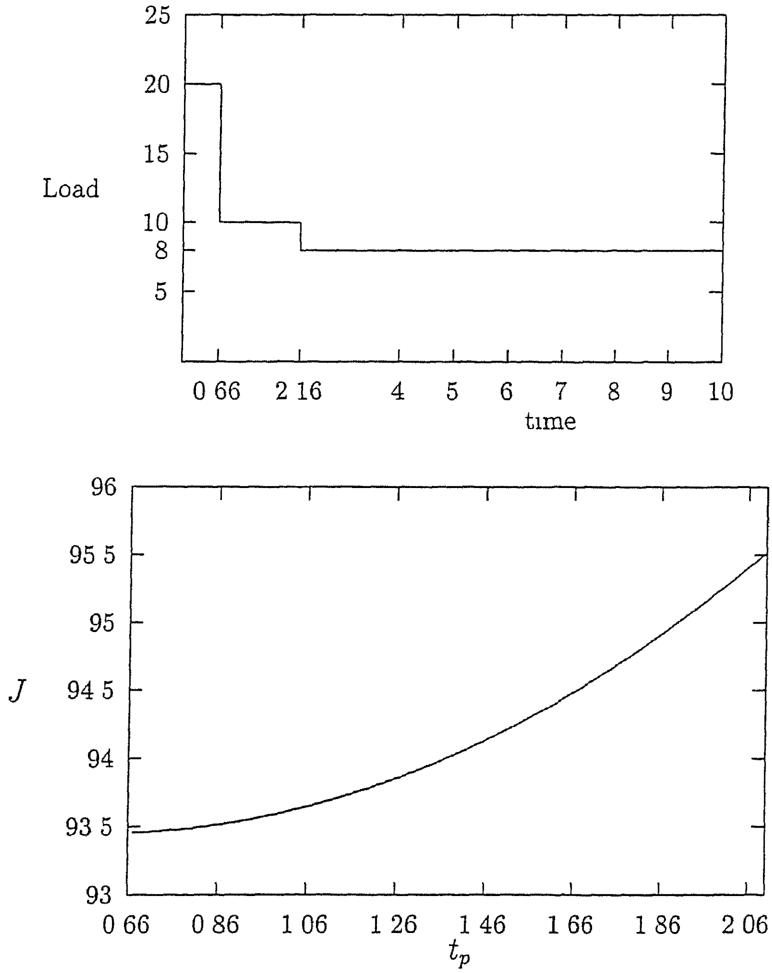

Figure 3 7 Performance vs. t_p for Example 2

The value of t_p at which the performance index is minimum = 0 67.

Let (t_i, t_i^*) be the interval following the i^{th} interval of partial routing in which $x_1(t) > x_{1s}$ and $x_1(t_i) = x_1(t_i^*) = x_{1s}$, for i=1,2,...n. Then we have,

$$\int_{t_i}^{t_i^*} \lambda(\tau).d\tau \quad = \quad C_1(t_i^* - t_i), \quad \text{for } i = 1, 2, .. \text{n} \qquad \text{(A)}$$

During the interval $[t_i^*, t_{s_{i+1}}]$, $\alpha_1(t) = 1$ and the network operation is in the linear mode. Therefore the dynamics of link 1 is given by $x_1 = -a_1 x_1 + \lambda(t)$ with the

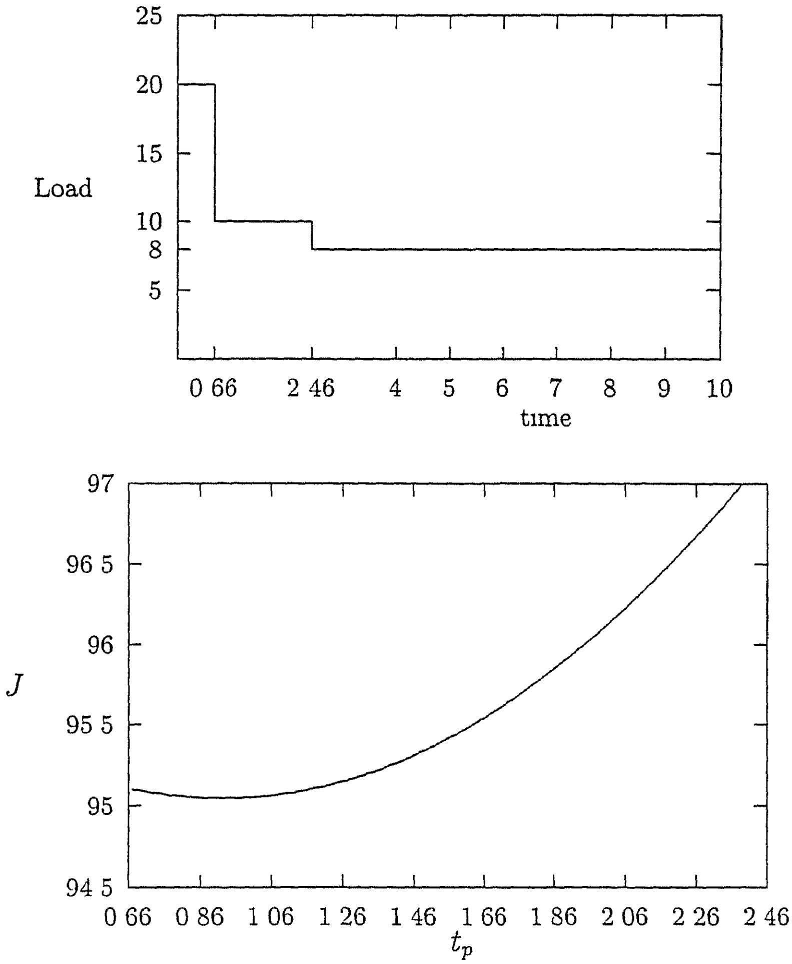

Figure 3 8: Performance vs. t_p for Example 3

The value of t_p at which the performance index is minimum= 0.916.

initial condition $x_1(t_i^*) = x_{1s}$. Therefore,

$x_1(t_{s_i+1}) = x_{1s}e^{-a_1(t_{s_i+1}-t_i^*)} + e^{-a_1 t_{s_i+1}} \int_{t_i^*}^{t_{s_i+1}} e^{a_1\tau}\lambda(\tau)d\tau$, i=1,.. (n-1)

Since $x_1(t_{s_i+1}) = x_{1s}$, we obtain the following (n-1) equations in t_{s_i}'s and t_i^*'s

$$x_{1s}[1 - e^{-a_1(t_{s_i+1}-t_i^*)}] = e^{-a_1\, t_{s_i+1}} \int_{t_i^*}^{t_{s_i+1}} e^{a_1\tau}\lambda(\tau)d\tau \qquad i = 1, 2, \,..\, (n-1) \qquad \text{(B)}$$

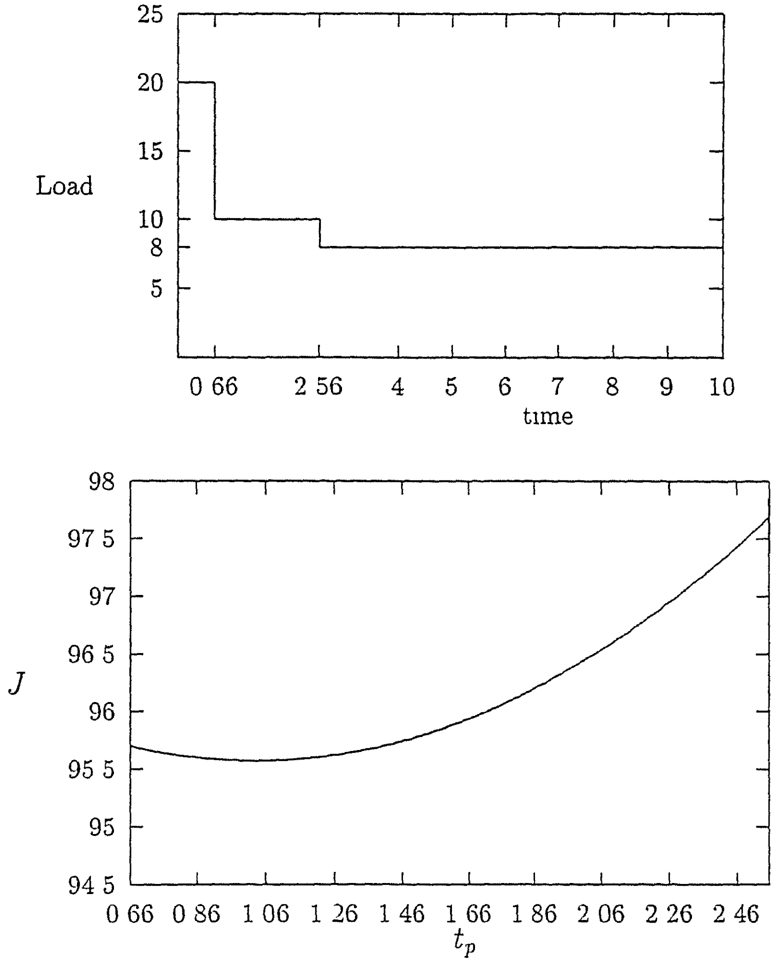

Figure 3 9: Performance vs. t_p for Example 4

The value of t_p at which the performance index is minimum = 1.06.

During the interval $[t_i, t_i^*]$, $x_1(t) > x_{1s}$, therefore $p_1 = -1$, and $p_1(t_i) = p_2(t_i) = \frac{(1-e^{a_2(t_i-T)})}{a_2}$ Therefore

$$\frac{(1-e^{a_2(t_i-T)})}{a_2} = (t_i^* - t_i) + p_1(t_i^*) \tag{3 18}$$

Over $[t_i^*, t_{s_i+1}]$, $p_1(t_i^*) = -1 + a_1 p_1$ and $p_1(t_{s_i+1}) = p_2(t_{s_i+1})$. Hence

$$p_1(t_i^*) = p_2(t_{s_i+1})e^{a_1(t_i^* - t_{s_i+1})} + \frac{(1 - e^{-a_1(t_{s_i+1} - t_i^*)})}{a_1}. \tag{3 19}$$

Substituting the r.h.s of the above in Equation (3 18), we obtain the following n equations.
For $i = 1, 2, \ldots, n$

$$\frac{1 - e^{a_2(t_i - T)}}{a_2} = (t_i^* - t_i) + p_2(t_{s_i+1})e^{a_1(t_i^* - t_{s_i+1})} + \frac{1 - e^{-a_1(t_{s_i+1} - t_i^*)}}{a_1} \tag{C}$$

To solve for $t_1, t_1^*, t_{s_2}, t_2, t_2^*, \ldots, t_{s_n}, t_n, t_n^*$ we have to solve the 3n-1 Equations (A), (B), and (C) given above.

3.5 Suboptimal Algorithm

From the discussion in the preceding section, the following conclusions can be drawn on the synthesis of the optimal routing strategy (ORS).

- To specify the ORS, the traffic pattern $\lambda(t)$ for the entire interval $[0, T]$, has to be known prior to the start of the network operation. An online implementation of the ORS is therefore not possible.

- As the number of positive crossings of the load pattern $\lambda(t)$ above the value equal to C_1 increases, the number of equations for specifying the intervals of partial routing increases by a factor 3. It is difficult to obtain analytical solutions to these equations and we expect that the numerical procedures employed to solve them would become more computationally intensive as n increases

The above reasons motivate the need for a suboptimal strategy Based on the practical considerations that link 1 can not carry packets any faster than its channel capacity C_1 (which is achieved when $x_1(t) = x_{1s}$) and once this maximum rate is achieved, the slower link (i e link 2) can be used for carrying the excess traffic

(than required to keep the flow on link 1 equal to C_1), we propose the following on-line implementable suboptimal algorithm

$$\alpha_1(t) = \begin{cases} 1 & ; \text{ over } [0, t_{s_1}] \\ \frac{C_1}{\lambda(t)} & ; \text{ over } (t_{s_i}, t_{b_i}), i = 1, 2, \quad n \\ 1 & ; \text{ elsewhere} \end{cases}$$

where t_{s_i}'s are the instants at which $x_1(t)$ reaches the value $=x_{1s}$ and t_{b_i}'s, as shown in Figure 3 5, are the instants corresponding to the negative crossings of the load pattern $\lambda(t)$ with the value equal to C_1

3.5.1 Numerical Examples

For a network with parameters $a_1 = 0\,9, a_2 = 0\,2, C_1 = 9, x_{1s} = 10, x_1(0) = 0, x_2(0) = 0, T = 10$ units we compare the performances of the optimal and suboptimal (proposed above) algorithms for the following load patterns

Case A:

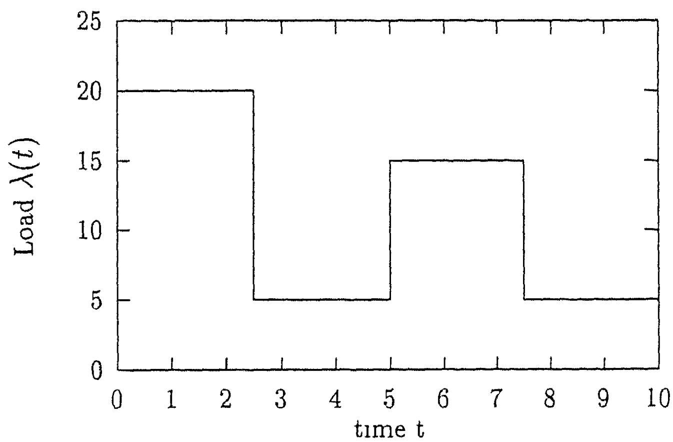

Table 3 1. Optimal and Suboptimal Strategies for Case A

Optimal Strategy	Suboptimal Strategy
$\alpha_1(t) = 1$ for $t \in [0, 0\ 6643]$	$\alpha_1(t) = 1$ for $t \in [0, 0\ 6643]$
$= 9/20$ for $t \in (0\ 6643, 1\ 900]$	$=9/20$ for $t \in (0\ 6643, 2.5]$
$=1$ for $t \in (1\ 9, 5\ 34]$	$=1$ for $t \in (2\ 5, 5\ 52]$
$=9/15$ for $t \in (5\ 34, 6\ 92]$	$=9/15$ for $t \in (5\ 52, 7\ 5)$
$=1$ for $t \in (6\ 92, 10]$.	$=1$ for $t \in [7\ 5, 10]$
$J_{opt} = 183\ 396$	$J_{subopt} = 194\ 317$

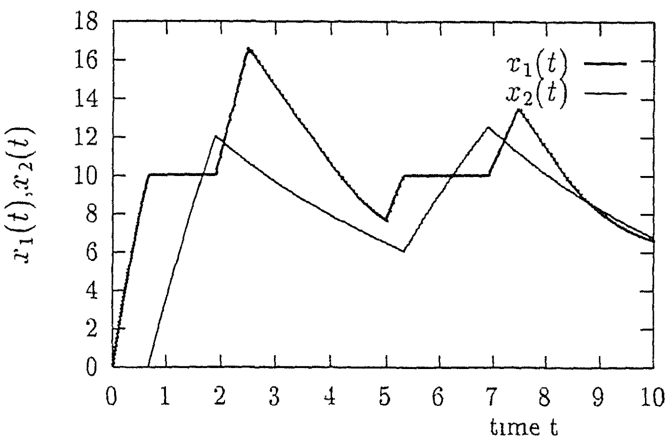

Figure 3.10. Buffer occupancies under the optimal strategy for Case A

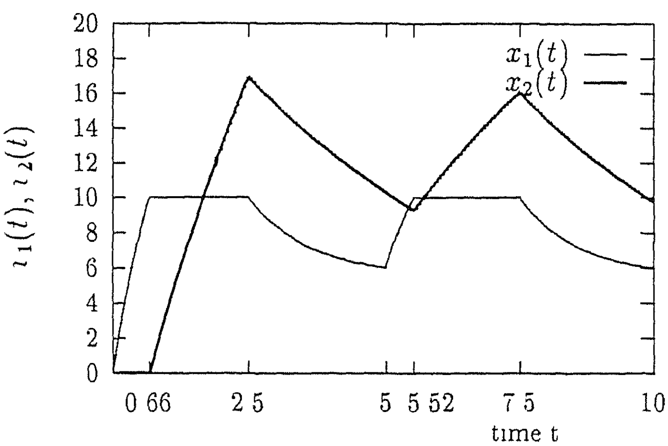

Figure 3 11 Buffer occupancies under the suboptimal strategy for Case A

Case B:

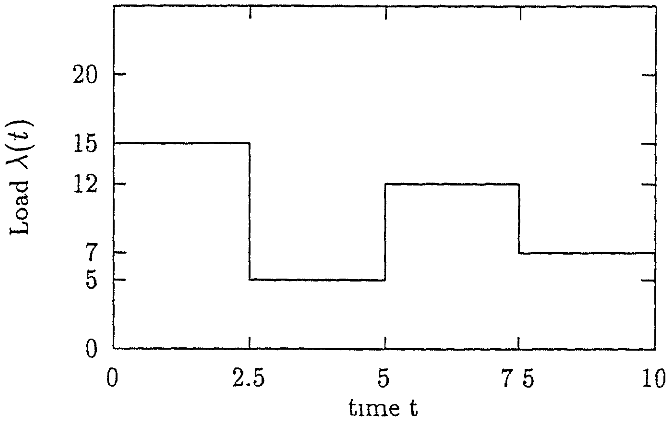

Table 3 2 Optimal and Suboptimal Strategies for Case B

Optimal Strategy	Suboptimal Strategy
$\alpha_1(t) = 1$ for $t \in [0, 1\ 0181]$ $= 9/15, t \in (1\ 0181, 1\ 480]$ $=1$ for $t \in (1\ 48, 5\ 6384]$ $=9/12$ for $t \in (5\ 6384, 6\ 93]$ $=1$ for $t \in (6\ 93, 10]$	$\alpha_1(t) = 1$ for $t \in [0, 1\ 0181]$ $=9/15, t \in (1\ 0181, 2\ 5]$ $=1, t \in (2\ 5, 5\ 8724]$ $=9/12, t \in (5\ 8724, 7\ 5)$ $=1$ for $t \in [7\ 5, 10]$
$J_{opt} = 122\ 2793$	$J_{subopt} = 131\ 7026$

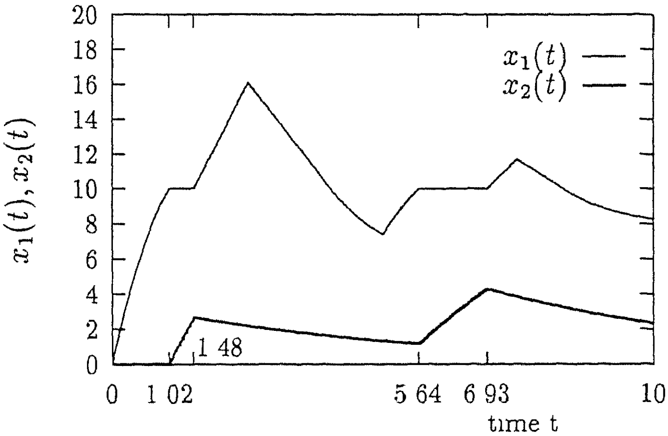

Figure 3 12. Buffer occupancies under the optimal strategy for Case B

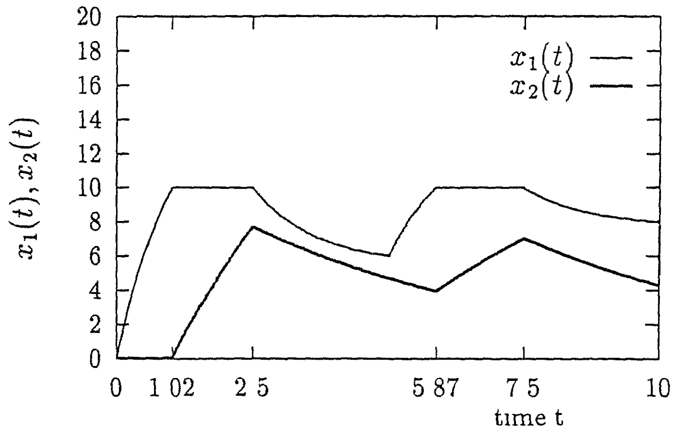

Figure 3 13 Buffer occupancies under the suboptimal strategy for Case B

Case C:

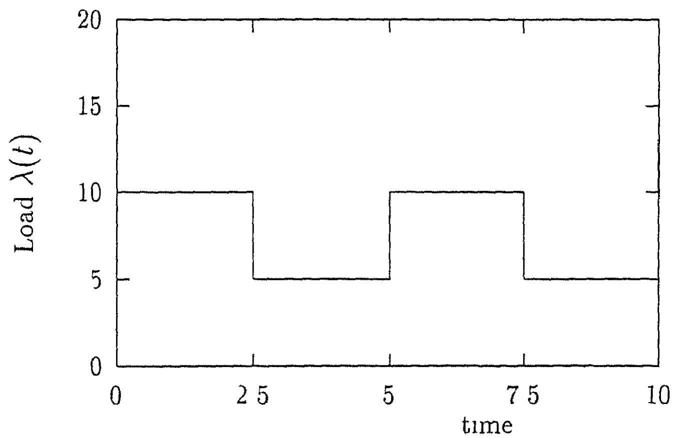

Table 3 3. Optimal and Suboptimal Strategies for Case C

Optimal Strategy	Suboptimal Strategy
$\alpha_1(t) = 1$ for $t \in [0, 10]$	$\alpha_1(t) = 1$ for $t \in [0, 6\ 6918]$ $= 9/10, t \in (6\ 6918, 7\ 5]$ $= 1, t \in (7\ 5, 10]$
$J_{opt} = 76\ 89797$	$J_{subopt} = 77\ 4746$

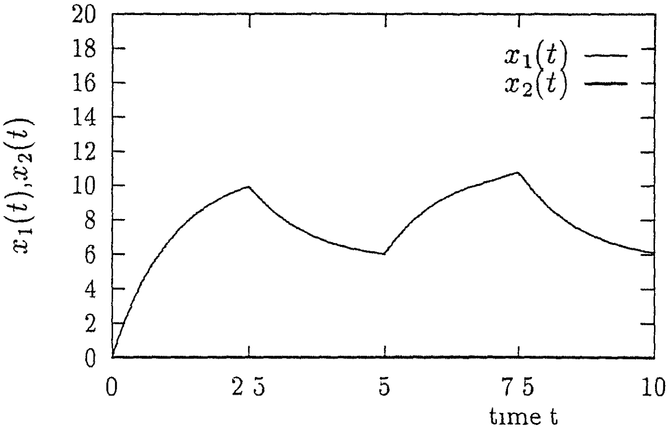

Figure 3 14 Buffer occupancies under the optimal strategy for Case C

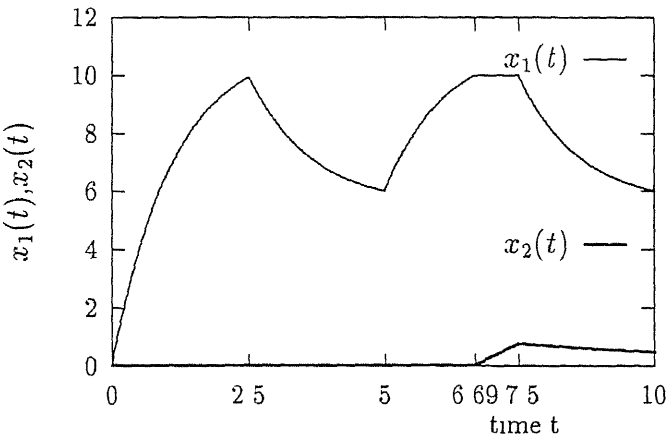

Figure 3 15: Buffer occupancies under the suboptimal strategy for Case C

Case D:

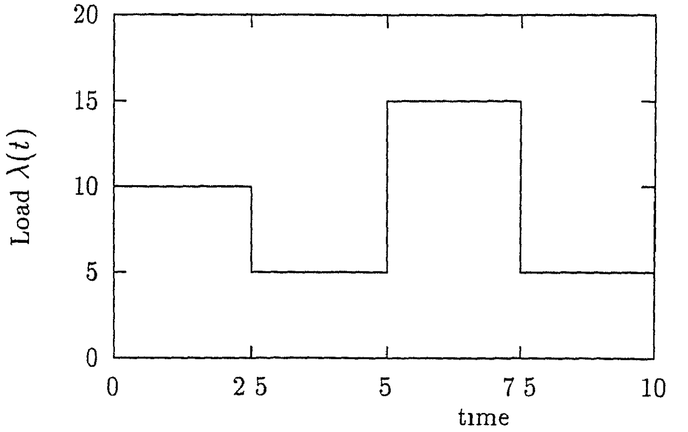

Table 3 4 Optimal and Suboptimal Strategies for Case D

Optimal Strategy	Suboptimal Strategy
$\alpha_1(t) = 1$ for $t \in [0, 5\ 520]$ $= 9/15$ for $t \in (5\ 52, 6\ 92]$ $= 1$ over $(6\ 92, 10]$	$\alpha_1(t) = 1$ for $t \in [0, 5\ 52]$ $=9/15, t \in (5\ 52, 7\ 5]$ $=1, t \in (7\ 5, 10]$
$J_{opt} = 105\ 25624$	$J_{subopt} = 106\ 9277$

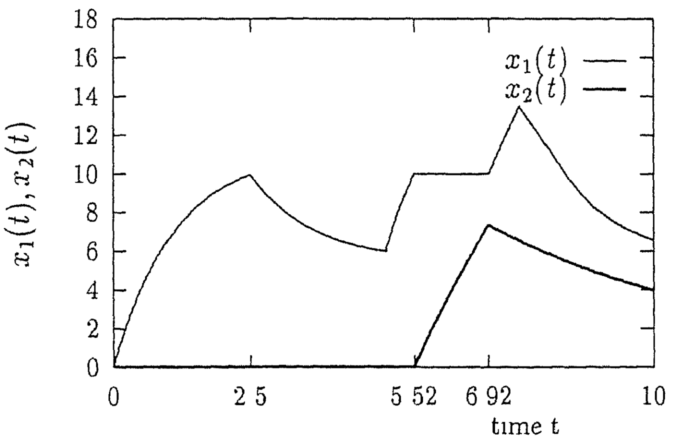

Figure 3 16 Buffer occupancies under the optimal strategy for Case D

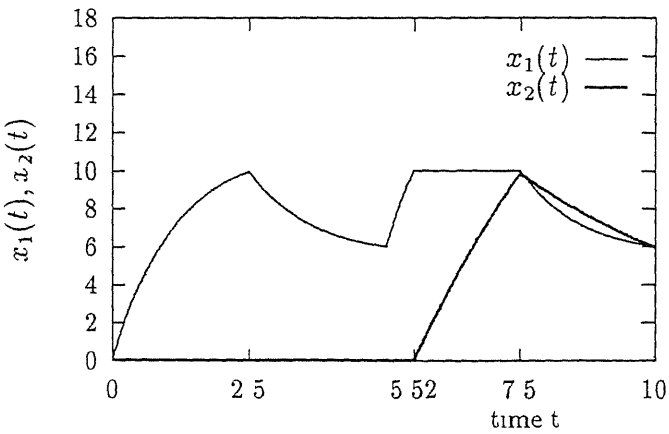

Fıgure 3.17 Buffer occupancıes under the suboptımal strategy for Case D

Case E:

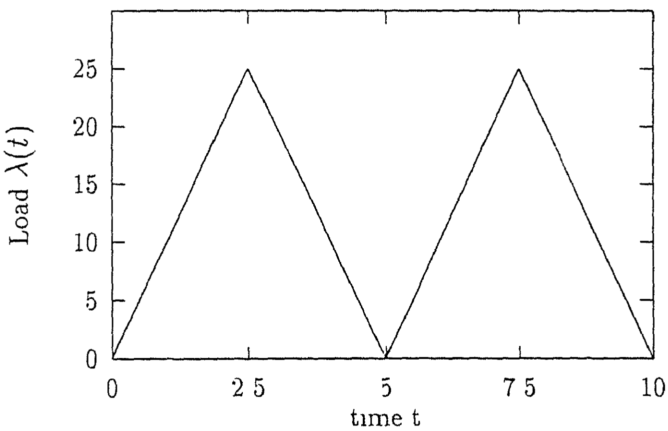

Table 3 5 Optimal and Suboptimal Strategies for Case E

Optimal Strategy	Suboptimal Strategy
$\alpha_1(t) = 1$ for $t \in [0, 1\ 79]$	$\alpha_1(t) = 1$ for $t \in [0, 1\ 79]$
$= 9/10t, t \in (1\ 79, 2\ 50]$	$=9/10t, t \in (1\ 79, 2\ 5]$
$= 9/(50 - 10t)$ for $t \in (2\ 5\ \ 3\ 21]$	$=9/(50\text{-}10t), t \in (2\ 5, 4\ 1]$
$=1$ for $t \in (3\ 21, 6\ 54]$	$=1, t \in (4\ 1, 6\ 61)$
$= 9/(10t - 50)$ for $t \in (6\ 54, 7\ 5]$	$= 9/(10t - 50)$ for $t \in [6\ 61, 7\ 5]$
$= 9/(100 - 10t)$ for $t \in (7\ 5\ \ 8\ 8]$	$=9/(100 - 10t)$ for $(7\ 5, 9\ 1]$
$=1$ over $(8\ 8, 10]$	$=1$ over $(9\ 1, 10]$
$J_{opt} = 209\ 2164$	$J_{subopt} = 213\ 50038$

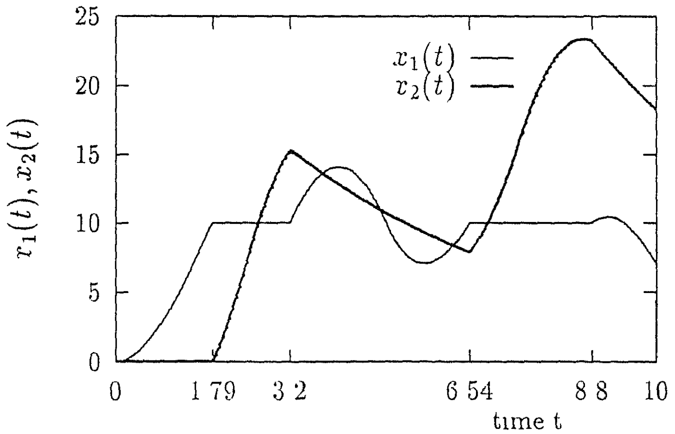

Figure 3 18: Buffer occupancies under the optimal strategy for Case E

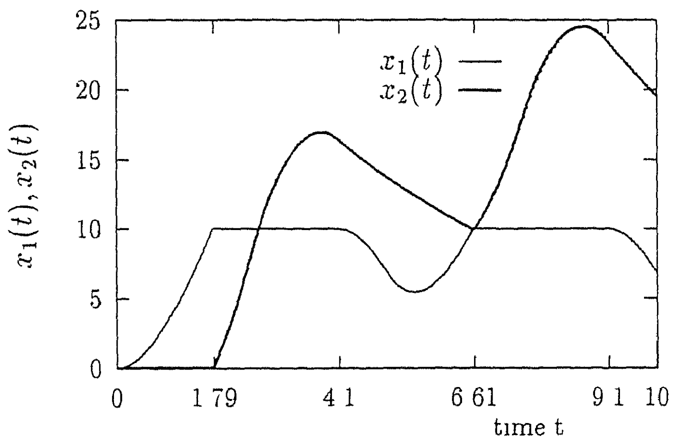

Figure 3.19: Buffer occupancies under the suboptimal strategy for Case E

Finally, we make the following observations regarding the nature of the optimal routing strategy for the case where the network operation starts in the saturation mode. If the initial buffer occupancy $x_1(0)$ is such that $x_1(0) > x_{1s} + C_1 T$, then $\alpha_1(t) = 0$ and $\alpha_2(t) = 1$, $\forall t \in [0\ T]$ is the optimal routing strategy. This follows from the fact the network operates in the saturation mode for the entire duration $[0, T]$ and therefore $\dot{p}_1 = -1$, $p_1(T) = 0$ Solution to this is $p_1(t) = T - t$ Since $T - t > p_2(t) = \frac{1-e^{a_2(t-T)}}{a_2}$, for all $t \in [0, T]$, $\alpha_1(t) \equiv 0$

The above result is in agreement with the following intuition also. Since the initial buffer occupancy $x_1(0)$ is large enough to guarantee that the link 1 can be operated at the maximum rate(C_1) for the entire duration $[0, T]$ even when no new packets are routed onto it, and since the rate at which packets are drawn out on link 2 increases linearly with its buffer occupancy, it is necessary to route all the traffic onto link 2 (for the entire duration) to minimize the total waiting time J

For initial buffer occupancies $x_1(0)$ such that $x_{1s} < x_1(0) < x_{1s} + C_1 T$, it can be argued (along the lines of Theorem 3 3 1) that the network operation can not end in saturation mode and hence ends either in the linear mode or in the transition mode It can also be argued that if $x_1(t)$ at any instant $t = t_0$ is equal to x_{1s} and if $\lambda(t)$ is greater than C_1 for all $t > t_0$, then $\alpha_1(t)$ equals $\frac{C_1}{\lambda(t)}$ is the optimal strategy in the interval $[t_0, T]$ An exhaustive investigation of these cases, is however, not attempted in this thesis

3.6 Conclusions

In this chapter we investigated the problem of optimal routing in a two node network in which the *faster* link has finite channel capacity, under the assumption that the buffer occupancies at the beginning of the network operation do not exceed the saturation values We argued that either the entire traffic $\lambda(t)$ is routed onto the *faster* link (i e α_1 equals unity in our case) or there is a partial routing of $\frac{C_1}{\lambda(t)}$ onto this link The equations required to be solved to specify the instants at which the network operation enters/comes out of, these *intervals of partial routing*, were

derived in terms of the link parameters a_1, a_2 and C_1 and the input traffic $\lambda(t)$ to the network A suboptimal algorithm which doesn't require the knowledge of the traffic pattern $\lambda(t)$ for the entire duration $[0, T]$, and, hence can be implemented on an on-line basis, was proposed and its performance compared with the optimal algorithm, in the case of some specific load patterns.

Chapter 4

Optimal and Suboptimal Routing Strategies for a Three Node Network

4.1 Introduction

In Chapter 1, we stated the motivations for the synthesis of optimal/suboptimal routing strategies for a large network from those for appropriately chosen, simpler network units which constitute it. With this perspective in the background, we investigated the problem of optimal traffic routing in a two-node network in the previous chapter. Under the assumptions that the faster link of this unit has a finite channel capacity and that the initial buffer occupancy for this link does not exceed its saturation value, we derived a set of equations (in terms of the link parameters of the network and the input traffic $\lambda(t)$) that are to be solved for specifying the optimal routing strategy. A suboptimal algorithm was also proposed for this network and its performance was compared with the optimal one in the case of some specific load patterns.

In this chapter, we investigate the problem of optimal routing in a basic network unit of three nodes shown in Figure 4.1. Nodes 1 and 2 of this network are source nodes which receive packets from the outside and node 3 is the destination node. A *direct path* and an *alternate(indirect) path* are provided at each source node. Link (1,3)

is the direct path at node 1, while the set of links {(1,2),(2,3)} constitute the alternate path Similarly link (2,3) is the direct path at node 2, while the alternate path is constituted of the set of links {(2,1),(1,3)} At the source nodes 1 and 2, dedicated buffers are assumed to be provided for each of their outgoing links Furthermore, we assume that the flowout on any link depends linearly on the buffer occupancy and has an upper bound equal to the channel capacity of that link

We first formulate the problem of optimal routing for this network unit in Section 4 2 of this chapter The properties and the implementation details of the optimal routing strategy for the case wherein all the links of the network have infinite channel capacities are discussed in Section 4 3 This corresponds to the linear mode of operation (of all the links) of the network It is shown that the optimal routing strategy depends only on the link parameters of the network and is independent of the input traffic $\lambda_1(t)$ and $\lambda_2(t)$ The optimal routing strategy is also shown to have the **loop-free** property

We then relax the above assumption (of infinite channel capacities for all the links) in Section 4 4 and investigate the case wherein one of the direct links (link (1,3) in this analysis) of the network is of finite channel capacity We show that the routing strategy at the source node 1 need not be *bang-bang* (in Section 4 5 we give examples wherein the routing variable $\alpha_{13}(t)$ takes non-zero, non-unity values during intervals of network operation) We prove that the optimal routing strategy has the **loop-free** property in this case also Of the 27 possible modes of network

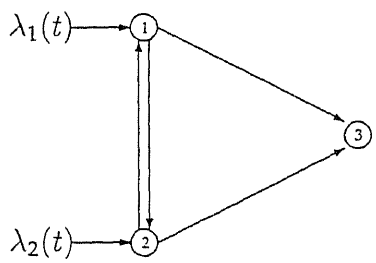

Figure 4 1. Network Topology of three nodes

operation (which are defined in Subsection 4 4 3) only 4 are possible in a terminal interval Numerical examples are cited to illustrate these cases.

In Section 4 5, we consider the case of networks with a direct link of finite channel capacity with certain (stated) assumptions on the link parameters a_{12}, a_{13}, a_{21} and a_{23} and on the initial buffer occupancies The implication of these assumptions (regarding the link parameters) is that, if all the links were of infinite capacity, then the optimal routing strategy would be direct routing for the entire duration of network operation. For the case wherein the initial buffer occupancy on the link with finite channel capacity is below the saturation value, we derive a set of equations (in terms of the link parameters and the input traffic), the solution to which specify the optimal routing strategy A suboptimal algorithm is also proposed and its performance is compared with the optimal one in the case of some illustrative examples

In Section 4 6, we consider the case wherein the flowout on a link is an exponential function of the (associated) buffer occupancy For small values of the buffer occupancy this model approximates the linear model, while for large values of the buffer occupancy the flowout reaches a saturation value equal to the channel capacity of that link The optimal routing strategy for this model is obtained by numerically solving the two point boundary value problem in the state and costate variables From the numerical investigations carried out for various choice of the link parameters, initial buffer occupancies and input traffic, we conjecture that the optimal routing strategy has the **loop-free** property in this case also. It is also observed that the network operation ends with a direct routing (at both the source nodes) of the input traffic under the optimal routing strategy. The state variables (buffer occupancies) and the costate variables under the optimal routing strategy are obtained numerically and they are shown in the form of graphs

Finally in Section 4 7, we recapitulate the major results in this chapter

4.2 Flow Model and the Problem Formulation

From the flow conservation principle and our assumed utilisation of link (j,k) in either the linear mode or the saturation mode, it is easily argued that the average number of packets x_{jk}, in the buffer associated with link (j,k) is governed by the equation

$$x_{jk} = \begin{cases} -a_{jk}x_{jk} + \lambda_{jk}^{in}(t) & \text{if } x_{jk} < C_{jk}/a_{jk} \quad \text{....(a)} \\ -C_{jk} + \lambda_{jk}^{in}(t) & \text{otherwise} \quad \text{.... (b)} \end{cases} \tag{4.1}$$

In the above, 4.1(a) corresponds to the linear mode of operation and 4.1(b) corresponds to the saturation mode of operation of link (j,k), $\lambda_{jk}^{in}(t)$ is the rate of input traffic flow to the link (j,k), C_{jk} is the channel capacity of the link (j,k) and the link parameter a_{jk} determines the rate at which packets are drawn out onto the link from the buffer.

It then follows that the dynamical evolution of the state of the network shown in Figure 4.1, when all its links operate in the linear mode, is given by

$$x_{12}(t) = -a_{12}x_{12}(t) + \alpha_{12}(t)[\lambda_1(t) + a_{21}x_{21}(t)] \tag{4.2}$$

$$x_{13}(t) = -a_{13}x_{13}(t) + \alpha_{13}(t)[\lambda_1(t) + a_{21}x_{21}(t)] \tag{4.3}$$

$$x_{21}(t) = -a_{21}x_{21}(t) + \alpha_{21}(t)[\lambda_2(t) + a_{12}x_{12}(t)] \tag{4.4}$$

$$x_{23}(t) = -a_{23}x_{23}(t) + \alpha_{23}(t)[\lambda_2(t) + a_{12}x_{12}(t)] \tag{4.5}$$

In the above set of equations $\alpha_{jk}(t)$ is defined as the fraction of the traffic routed on link (j,k) from the total traffic arriving at node j Since there is no accumulation of traffic at the nodes, we have the normalizing conditions,

$\alpha_{12}(t) + \alpha_{13}(t) = 1$, $\alpha_{21}(t) + \alpha_{23}(t) = 1$, $0 \leq \alpha_{jk}(t) \leq 1$ for all (j,k).

Whenever for any link (j,k), the expression $a_{jk}x_{jk}(t)$ is greater than C_{jk}, that link goes into saturation and the term $a_{jk}x_{jk}(t)$ is to be replaced by the constant C_{jk}

The problem of synthesising the optimal routing strategy for this network can be stated as follows:

Given the link parameters a_{12}, a_{13}, a_{21}, a_{23} and the channel capacities C_{12}, C_{13}, C_{21} and C_{23} of the network whose system dynamics is given by the Equations (4.2),

(4.3), (4.4) and (4.5) find the routing variables $\alpha_{12}(t)$, $\alpha_{13}(t)$, $\alpha_{21}(t)$ and $\alpha_{23}(t)$ which satisfy the normalizing conditions, so as to minimise the performance index

$$J = \int_0^T (x_{12} + x_{13} + x_{21} + x_{23})dt \tag{4.6}$$

of the network for the time span $[0, T]$ of operation.

The rationale behind choosing the above criterion J as the performance index is that a waiting cost is incurred at a rate proportional to the number of customers in the system.

4.3 Analysis of the Linear Mode

Assumption 4.3.1 We assume in this section that all the links of the network have infinite channel capacities. Therefore the network operates in the linear mode for any input traffic $\lambda_1(t)$ and $\lambda_2(t)$.

We find the optimal routing variables by applying Pontryagin's Maximum principle [22,23] to the system (4.2), (4.3), (4.4) and (4.5). The Hamiltonian H for the system is given by $H = \sum_{(j,k)} x_{jk} + \underline{p}^T \underline{x}$ where

$$\underline{p} = \begin{bmatrix} p_{12}(t) \\ p_{13}(t) \\ p_{21}(t) \\ p_{23}(t) \end{bmatrix}$$

and

$$\underline{x} = \begin{bmatrix} x_{12}(t) \\ x_{13}(t) \\ x_{21}(t) \\ x_{23}(t) \end{bmatrix}$$

In the above, $p_{12}(t)$, $p_{13}(t)$, $p_{21}(t)$ and $p_{23}(t)$ are the costate variables.

According to the maximum principle, the optimal state and costate variables satisfy the following necessary conditions.

(i) $\dot{p}_{jk} = -\partial H/\partial x_{jk}$ with $p_{jk}(T) = 0$

(ii) $\dot{x}_{jk} = \partial H/\partial p_{jk}$.

(iii) On the optimal trajectory, we have $H_{opt} = \min_{\alpha_{jk}} H(x_{jk}, p_{jk}, \alpha_{jk})$.

It is easily argued that minimising H w.r.t $\alpha_{jk}(t)$'s is equivalent to minimising the expression, $\alpha_{12}(p_{12} - p_{13})(\lambda_1 + a_{21}x_{21}) + \alpha_{21}(p_{21} - p_{23})(\lambda_2 + a_{12}x_{12})$. Here, the terms $(\lambda_1 + a_{21}x_{21})$ and $(\lambda_2 + a_{12}x_{12})$ are non-negative. And as the non-negative routing variables α_{12} and α_{21} are bounded by a value equal to unity, the optimal routing strategy is the *bang-bang* type of control [23] with the following possible regimes of operation.

Regime 1 : $p_{13}(t) > p_{12}(t)$ and $p_{23}(t) > p_{21}(t)$ with $\alpha_{12}(t) \equiv \alpha_{21}(t) \equiv 1$ for all t in an interval that is a subset of [0,T].

Regime 2 : $p_{13}(t) > p_{12}(t)$ and $p_{23}(t) < p_{21}(t)$ with $\alpha_{12}(t) \equiv 1$ and $\alpha_{21}(t) \equiv 0$ for all t in an interval that is a subset of [0,T]

Regime 3 : $p_{13}(t) < p_{12}(t)$ and $p_{23}(t) > p_{21}(t)$ with $\alpha_{12}(t) \equiv 0$, $\alpha_{21}(t) \equiv 1$ for all t in an interval that is a subset of [0,T]

Regime 4 : $p_{13}(t) < p_{12}(t)$ and $p_{23}(t) < p_{21}(t)$ with $\alpha_{12}(t) \equiv 0$, $\alpha_{21}(t) \equiv 0$ for all t in an interval that is a subset of [0,T]

The costate variables $p_{12}(t)$, $p_{21}(t)$, $p_{13}(t)$ and $p_{23}(t)$ are given as the solutions to the following differential equations

$$\dot{p}_{12}(t) = -1 + a_{12}p_{12}(t) - a_{12}p_{21}(t)\alpha_{21}(t) - a_{12}p_{23}(t)(1 - \alpha_{21}(t)) \tag{4.7}$$

$$\dot{p}_{21}(t) = -1 + a_{21}p_{21}(t) - a_{21}p_{12}(t)\alpha_{12}(t) - a_{21}p_{13}(t)(1 - \alpha_{12}(t)) \tag{4.8}$$

$$\dot{p}_{13}(t) = -1 + a_{13}p_{13}(t) \tag{4.9}$$

$$\dot{p}_{23}(t) = -1 + a_{23}p_{23}(t) \tag{4 10}$$

Equations (4.9) and (4 10) can be solved independently to yield

$$p_{13}(t) = \frac{1 - e^{a_{13}(t-T)}}{a_{13}} \tag{4 11}$$

$$p_{23}(t) = \frac{1 - e^{a_{23}(t-T)}}{a_{23}} \tag{4.12}$$

4.3.1 Optimal Solution and its Properties

Towards determining the other optimal costate trajectories and switchings between the various regimes of operation we state below theorems about the conditions in which link (1,2) and/or (2,1) can or cannot carry any packets.

Theorem 4.3.1 *In the linear mode of operation of all the links of the network, the optimal routing strategy is such that incoming packets are not routed on links(1,2) and link (2,1) simultaneously over an interval* $[t_0, T]$ *with* $t_0 \in [0, T]$

Proof : We prove this by contradiction Accordingly let $\alpha_{12}(t) \equiv \alpha_{21}(t) \equiv 1$, for all t in an interval $[t_0, T]$ This corresponds to regime 1 with $p_{13}(t) > p_{12}(t)$ and $p_{23}(t) > p_{21}(t)$, for all t in $[t_o, T]$ But Equations (4 7) and (4 8) with $\alpha_{12}(t) \equiv \alpha_{21}(t) \equiv 1$ reduce to

$$p_{12} = -1 + a_{12}(p_{12} - p_{21}) \tag{4.13}$$

$$p_{21} = -1 + a_{21}(p_{21} - p_{12}) \tag{4.14}$$

with the boundary conditions $p_{12}(T) = p_{21}(T) = 0$. The solutions to the above Equations (4 13) and (4 14) are

$$p_{12}(t) = p_{21}(t) = T - t, \quad \forall t \in [t_0, T]$$

$$\text{But } \ (T-t) > \frac{1 - e^{a_{13}(t-T)}}{a_{13}} \quad \forall t \in [t_o, T) \text{ and for } a_{13} > 0$$

Therefore $p_{12}(t) > p_{13}(t), \forall t \in [t_o, T)$ Similarly $p_{21}(t) > p_{23}(t), \forall t \in [t_0, T)$

The last two inequalities contradict the hypothesis that the system is operating in regime 1 over the interval $[t_0, T]$ This concludes the proof

□

Theorem 4 3 1 simply asserts that the optimal routing strategy forbids operation that ends in regime 1

Having seen that the optimal routing strategy is one in which incoming packets are not simultaneously routed over links (1,2) and (2,1) in any interval that ends in

T, we now state the conditions under which the incoming packets are never routed on these links for any interval that is a subset of $[0, T]$.

Theorem 4.3.2 *In the linear mode of operation of all the links, if the link parameters satisfy any of the following sets of conditions,*

(i) $a_{23} > a_{13} > a_{12}$

(ii) $a_{13} > a_{23} > a_{21}$

(iii) $a_{13} > a_{12}$ *and* $a_{23} > a_{21}$

then the system, under the optimal routing strategy, operates in regime 4 for the entire duration $[0, T]$

Proof : We shall first show that under any of the three sets of conditions, the system has to end in regime 4 We prove this by a contradiction as follows Consider that the system is operating in regime 2 in an interval $[t_o, T]$ The equations for $p_{12}(t)$ and $p_{21}(t)$ are given as

$$\begin{aligned} p_{12}(t) &= -1 + a_{12}p_{12}(t) - a_{12}p_{23}(t) && (4\ 15) \\ p_{21}(t) &= -1 + a_{21}p_{21}(t) - a_{21}p_{12}(t) && (4\ 16) \end{aligned}$$

with $p_{12}(T) = p_{21}(T) = 0$
Using (4.12), Equation (4 15) can be solved to obtain, $\forall t \in [t_0, T]$,

$$p_{12}(t) = \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{(a_{12} - a_{23})} \qquad (4\ 17)$$

Denoting by $q(t)$ the last two terms in Equation (4.17) i.e.

$$q(t) = \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{23}(t-T)} - e^{a_{12}(t-T)}}{a_{23} - a_{12}} \qquad (4\ 18)$$

Observe that $q(t)$ (defined above) is non-negative and monotonically decreasing over $[0, T]$ with $q(T) = 0$ since

$$q(t) = \begin{cases} a_{12}\left\{\frac{e^{a_{23}(t-T)} - e^{a_{12}(t-T)}}{a_{23} - a_{12}}\right\} & < 0\ , \forall t \in [0, T) \\ 0 & \text{at t = T} \end{cases}$$

$$\text{Hence } p_{12}(t) \geq \frac{1 - e^{a_{12}(t-T)}}{a_{12}}, \quad \forall t \in [t_o, T]$$

For the system to operate in regime 2, we must have

$$p_{13}(t) = \frac{1 - e^{a_{13}(t-T)}}{a_{13}} > p_{12}(t)$$

Therefore, we must have

$$\frac{1 - e^{a_{13}(t-T)}}{a_{13}} > \frac{1 - e^{a_{12}(t-T)}}{a_{12}}, \quad \forall t \in [t_o, T),$$

which implies that the system could possibly end in regime 2 only if $a_{12} > a_{13}$ Further, as $p_{12}(t)$ given in (4 17) is symmetric w r t a_{12} and a_{23}, by a similar argument it also follows that regime 2 could possibly be the terminal regime only if $a_{23} > a_{13}$ Under any of the three sets of conditions, one of these inequalities ($a_{12} > a_{13}$ and $a_{23} > a_{13}$) is violated and therefore the system cannot end in regime 2

An analogous reasoning on the nature of $p_{21}(t)$ and $p_{23}(t)$ dynamics of regime 3, leads to the result that the system could possibly end in regime 3 only if $a_{13} > a_{23}$ and $a_{21} > a_{23}$ Under any of the three sets of conditions, one of these inequalities is always violated and therefore the system cannot end in regime 3. Since by Theorem 4 3 1, the system can not end in regime 1, it follows that under the conditions (i), (ii) and (iii) the terminal regime is regime 4.

Corresponding to the terminal regime 4, the dynamics of $p_{12}(t)$ and $p_{21}(t)$ are as follows

$$p_{12}(t) = \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{(a_{12} - a_{23})}$$

and

$$p_{21}(t) = \frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{(a_{21} - a_{13})}$$

It is easily argued that (the argument is the same as used in the initial part) under any of the three sets of conditions, the dynamics of $p_{12}(t)$, $p_{13}(t)$, $p_{21}(t)$ and $p_{23}(t)$ of regime 4 are such that

$$p_{12}(t) > p_{13}(t), \quad \forall t \in [0, T)$$

$$p_{21}(t) > p_{23}(t), \quad \forall t \in [0, T)$$

Hence the system operates in regime 4 for the entire duration $[0, T]$ under any of the three sets of conditions (i), (ii), and (iii). This concludes the proof.

□

Now we show that the system has to end in regime 4, regardless of the non-negative values of a_{ij}.

Theorem 4.3.3 *In the linear mode of operation of all the links of the network, under the optimal routing strategy, the system always ends in regime 4*

Proof : The following statements were proved in the proof of Theorem 4.3 2

(i) The system possibly ends in regime 2 only if $a_{12} > a_{13}$ and $a_{23} > a_{13}$

(ii) The system possibly ends in regime 3 only if $a_{21} > a_{23}$ and $a_{13} > a_{23}$

Assume that $a_{12} > a_{13}$ and $a_{23} > a_{13}$ Under this condition the possibility of regime 3 being the terminal regime is ruled out By Theorem 4 3.1, regime 1 can never be the terminal regime Therefore the system ends either in regime 2 or regime 4 But for both these regimes, dynamics of $p_{12}(t)$ is the same

$$p_{12}(t) = \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{a_{12} - a_{23}}$$

$$\text{As } a_{23} > a_{13}, \quad \frac{1 - e^{a_{13}(t-T)}}{a_{13}} > \frac{1 - e^{a_{23}(t-T)}}{a_{23}}$$

$$\text{and as } a_{12} > a_{13}, \quad \frac{1 - e^{a_{13}(t-T)}}{a_{13}} > \frac{1 - e^{a_{12}(tT)}}{a_{12}}, \ \forall t \in (-\infty, T)$$

$$\begin{aligned} \text{Let } \frac{1 - e^{a_{13}(t-T)}}{a_{13}} &= \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + f_{23}(t), \ \text{ where } f_{23}(t) > 0, \forall t \in (-\infty . T) \\ &= \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + f_{12}(t), \ \text{ where } f_{12}(t) > 0, \forall t \in (-\infty \ T) \end{aligned}$$

Then, it can be shown that,

$$\begin{aligned} p_{12}(t) &= \frac{1 - e^{a_{13}(t-T)}}{a_{13}} + \frac{a_{23} f_{12}(t) - a_{12} f_{23}(t)}{a_{12} - a_{23}} \\ &= p_{13}(t) + \frac{a_{23} f_{12}(t) - a_{12} f_{23}(t)}{a_{12} - a_{23}} \end{aligned}$$

It can be proved, as follows, that over an interval $I = [t_o, T)$, for some $t_o < T$, the function $p_{12}(t)$ is greater than $p_{13}(t)$ In other words, in the left vicinity of $t = T$, $p_{12}(t) > p_{13}(t)$

$$p_{12}(t) > p_{13}(t) \text{ iff } \frac{a_{23}f_{12}(t) - a_{12}f_{23}(t)}{a_{12} - a_{23}} > 0$$

$$\text{Let } a_{12} > a_{23} \text{ Then } p_{12}(t) > p_{13}(t) \text{ iff } \frac{f_{12}(t)}{f_{23}(t)} > \frac{a_{12}}{a_{23}}$$

For t in the vicinity of T, using Taylor series expansion around T,

$$\begin{aligned}\frac{f_{12}(t)}{f_{23}(t)} &= \frac{\left(\frac{1 - e^{a_{13}(t-T)}}{a_{13}}\right) - \left(\frac{1 - e^{a_{12}(t-T)}}{a_{12}}\right)}{\left(\frac{1 - e^{a_{13}(t-T)}}{a_{13}}\right) - \left(\frac{1 - e^{a_{23}(t-T)}}{a_{23}}\right)} \\ &\approx \frac{a_{12} - a_{13}}{a_{23} - a_{13}},\end{aligned}$$

after neglecting terms containing $(t - T)$ and its higher orders

$$\text{Since } \frac{a_{12} - a_{13}}{a_{23} - a_{13}} > \frac{a_{12}}{a_{23}},$$

it follows that $p_{12}(t) > p_{13}(t)$ in some interval $[t_o, T)$ (The argument for the case $a_{23} > a_{12}$ is along similar lines as above) Thus the expression for $p_{12}(t)$ and $p_{13}(t)$ in both regimes 2 and 4 are such that $p_{12}(t) > p_{13}(t)$ for some interval ending with T Therefore it follows that the system can never end in regime 2. By a similar argument as above, concerning the dynamics of $p_{21}(t)$ and $p_{23}(t)$, it can be concluded that the system never ends in regime 3 Hence it follows that the system, under the optimal routing strategy always ends in regime 4, regardless of the non-negative values of the link parameters a_{jk}. This concludes the proof

□

Theorem 4.3.4 *In the linear mode of operation of all the links of the network, under the optimal routing strategy, the terminal regime 4 is never preceded by regime 1*

A necessary and sufficient condition for the terminal regime 4 to be preceded by regime 2 is

$$\frac{1 - e^{-a_{12}T}}{a_{12}} + \frac{1 - e^{-a_{23}T}}{a_{23}} + \frac{e^{-a_{12}T} - e^{-a_{23}T}}{a_{12} - a_{23}} < \frac{1 - e^{-a_{13}T}}{a_{13}}$$

which for large enough T, reduces to the inequality

$$\frac{1}{a_{12}} + \frac{1}{a_{23}} < \frac{1}{a_{13}}$$

A necessary and sufficient condition for the terminal regime 4 to be preceded by regime 3 is

$$\frac{1-e^{-a_{21}T}}{a_{21}} + \frac{1-e^{-a_{13}T}}{a_{13}} + \frac{e^{-a_{21}T}-e^{-a_{13}T}}{a_{21}-a_{13}} \quad < \quad \frac{1-e^{-a_{23}T}}{a_{23}},$$

which for large enough T, reduces to the inequality

$$\frac{1}{a_{21}} + \frac{1}{a_{13}} < \frac{1}{a_{23}}.$$

Proof: The dynamics of $p_{12}(t), p_{21}(t), p_{13}(t)$ and $p_{23}(t)$ in the terminal regime 4 are given as the following:

$$\begin{aligned}
p_{12}(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)}-e^{a_{23}(t-T)}}{a_{12}-a_{23}}, \\
p_{21}(t) &= \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)}-e^{a_{13}(t-T)}}{a_{21}-a_{13}}, \\
p_{13}(t) &= \frac{1-e^{a_{13}(t-T)}}{a_{13}}, \\
p_{23}(t) &= \frac{1-e^{a_{23}(t-T)}}{a_{23}}
\end{aligned}$$

As shown in the proof of Theorem 4 3 2, if $a_{13} \geq a_{23}$ then $p_{12}(t)$ as given by the above expression will be greater than $p_{13}(t)$ in the entire interval $[0, T)$ But if $a_{23} > a_{13}$, then $p_{21}(t) > p_{23}(t)$ $\forall t \in [0, T)$ Since either of these two is always the case, it follows that either $p_{12}(t) > p_{13}(t)$ or $p_{21}(t) > p_{23}(t)$ (or possibly both) for the entire duration $[0\ T)$ For the terminal regime 4 to be preceded by regime 1, there must exist some $t_o \in [0, T]$ when $p_{12}(t_o) = p_{13}(t_o)$ and $p_{21}(t_o) = p_{23}(t_o)$ Since this cannot be satisfied for any t_o, it follows that the terminal regime 4 can never be preceded by regime 1. It also follows that a necessary and sufficient condition for the terminal regime 4 (in which $p_{12}(t) > p_{13}(t)$) to be preceded by regime 2 (in which $p_{12}(t) < p_{13}(t)$) is $p_{12}(0) < p_{13}(0)$.

i.e

$$\frac{1-e^{-a_{12}T}}{a_{12}} + \frac{1-e^{-a_{23}T}}{a_{23}} + \frac{e^{-a_{12}T}-e^{-a_{23}T}}{a_{12}-a_{23}} \quad < \quad \frac{1-e^{-a_{13}T}}{a_{13}},$$

which for large enough T, reduces to the following

$$\frac{1}{a_{12}} + \frac{1}{a_{23}} < \frac{1}{a_{13}}.$$

Similarly, a necessary and sufficient condition for the terminal regime 4 to be preceded by regime 3 is $p_{21}(0) < p_{23}(0)$

i.e

$$\frac{1-e^{-a_{21}T}}{a_{21}} + \frac{1-e^{-a_{13}T}}{a_{13}} + \frac{e^{-a_{21}T} - e^{-a_{13}T}}{a_{21} - a_{13}} \quad < \quad \frac{1-e^{-a_{23}T}}{a_{23}},$$

which for large enough T, reduces to the inequality·

$$\frac{1}{a_{21}} + \frac{1}{a_{13}} < \frac{1}{a_{23}}$$

This concludes the proof of Theorem 4 3 4.

□

Towards proving Theorem 4 3 5 we make use of the following lemma.

Lemma 4.3.1 *For any choice of the parameters a_{12}, a_{13}, a_{23} such that $a_{12} > 0$, $a_{13} > 0$ and $a_{23} > 0$, the functions*

$$\begin{aligned} p_{12}(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{a_{12} - a_{23}} \\ \textit{and} \quad p_{13}(t) &= \frac{1-e^{a_{13}(t-T)}}{a_{13}} \end{aligned}$$

can have at the most one point of intersection in the interval $(-\infty, T)$

We make use of the following result by Pontryagin [22] stated below without proof towards proving the above lemma.

Lemma 4.3.2 *Let λ_1, λ_2, $,\lambda_m$ be distinct real numbers, and let $f_1(t)$, $f_2(t)$, . . $,f_m(t)$ be polynomials (with real coefficients) of degree k_1, k_2, . . $.k_m$ respectively Then, the functions $f_1(t)e^{\lambda_1 t} + f_2(t) + \quad + f_m e^{\lambda_m t}$ has at most $k_1 + k_2 + \;\; . \;\; + k_m + (m-1)$ roots*

Proof : Let us assume that $p_{12}(t)$ and $p_{13}(t)$ have two points of intersection in $(-\infty, T)$ (denoted as t_1 and t_2 where $t_1 < t_2$) In other words the function $(p_{12}(t) - p_{13}(t))$ is zero at $t = t_1$ and at $t = t_2$ Note that $(p_{12}(T) - p_{13}(T))$ is also zero From Rolles Theorem, it can be concluded that the derivative of the function $(p_{12}(t) - p_{13}(t))$ is zero at least once in the open intervals (t_1, t_2) and (t_2, T). Thus the function $(p_{12} - p_{13})$ has at least two roots in the interval $(-\infty, T)$.
The expression for $(p_{12}(t) - p_{13}(t))$ is given as

$$p_{12}(t) - p_{13}(t) = \frac{a_{23}e^{a_{12}(t-T)}}{(a_{12} - a_{23})} - \frac{a_{12}e^{a_{23}(t-T)}}{(a_{12} - a_{13})} + e^{a_{13}(t-T)}$$

At $t = T$, $p_{12} - p_{13} = 0$
Thus the function $p_{12} - p_{13}$ has at least three roots in the interval $(-\infty, T]$ But by Lemma 4 3.2, this function (which is a sum of three exponential terms weighted by constants) can have at most two roots in $(-\infty, \infty)$. Thus the assumption that $p_{12}(t)$ and $p_{13}(t)$ have two points of intersection in the interval $(-\infty, T)$ leads to a contradiction
The above arguments can be easily extended to prove that the assumption $p_{12}(t)$ and $p_{13}(t)$ has n (where $n > 2$) points of intersection leads to the conclusion that the function $(p_{12} - p_{13})$ has at least $n + 1$ roots. Since this can not be the case, it follows that $p_{12}(t)$ and $p_{13}(t)$ can not have two or more points of intersection in the interval $(-\infty, T)$ Hence the lemma.

□

Assuming that the terminal regime 4 is preceded by regime 2 (regime 3), we now show that such a regime 2 (regime 3) couldn't have been preceded by either regime 1 or regime 4 or regime 3 (regime 2)

Theorem 4.3.5 *The system, under the linear mode of operation of all the links with the optimal routing strategy, can not end with any of the following sequence of regimes*

(a) *regime 1→regime 2→regime 4.*

(b) *regime 3→regime 2→regime 4*

(c) regime 4→regime 2→regime 4

(d) regime 1→regime 3→regime 4

(e) regime 2→regime 3→regime 4.

(f) regime 4→regime 3→regime 4

Proof : (a) We prove this by contradiction as follows:
Let $I_1 = [t_o, t_1), I_2 = [t_1, t_2)$ and $I_4 = (t_2, T]$ be the intervals corresponding to regime 1, regime 2 and regime 4 (modes of operation) respectively. During $I_1, p_{12}(t) < p_{13}(t)$ and $p_{21}(t) < p_{23}(t)$ and

$$p_{12}(t) = -1 + a_{12}p_{12}(t) - a_{12}p_{21}(t) \tag{4.19}$$

$$p_{21}(t) = -1 + a_{21}p_{21}(t) - a_{21}p_{12}(t) \tag{4.20}$$

The solution to the above, in terms of the initial values $p_{12}(t_o)$ and $p_{21}(t_o)$ are given as

$$p_{12}(t) = t_o - t + \frac{a_{21}p_{12}(t_o) + a_{12}p_{21}(t_o)}{a_{12} + a_{21}} + \frac{a_{12}(p_{12}(t_o) - p_{21}(t_o))e^{(a_{12}+a_{21})(t-t_o)}}{a_{12} + a_{21}} \tag{4.21}$$

$$p_{21}(t) = t_o - t + \frac{a_{12}p_{21}(t_o) + a_{21}p_{12}(t_o)}{a_{12} + a_{21}} - \frac{a_{21}(p_{12}(t_o) - p_{21}(t_o))e^{(a_{12}+a_{21})(t-t_o)}}{a_{12} + a_{21}} \tag{4.22}$$

The slopes of $p_{12}(t)$ and $p_{21}(t)$ are given as

$$\begin{aligned} p_{12}(t) &= -1 + a_{12}[p_{12}(t_o) - p_{21}(t_o)]e^{(a_{12}+a_{21})(t-t_o)} \\ p_{21}(t) &= -1 + a_{21}[p_{21}(t_o) - p_{12}(t_o)]e^{(a_{12}+a_{21})(t-t_o)} \end{aligned}$$

Since $p_{13}(t) = \dfrac{1 - e^{a_{13}(t-T)}}{a_{13}}$ and $p_{23}(t) = \dfrac{1 - e^{a_{23}(t-T)}}{a_{23}}$

$$\begin{aligned} p_{13}(t) &= -e^{a_{13}(t-T)} \\ p_{23}(t) &= -e^{a_{23}(t-T)} \end{aligned}$$

□

The implications of the Theorems 4.3.1 to 4.3.5 can be summarised as the following: In the linear mode of operation of all the links, under the optimal routing strategy,

- The system always ends in regime 4, in which the direct links(1,3) and (2,3) are used.
- It has at the most one inter-regime switching in $[0, T]$
- The system never enters regime 1 in any interval that is a subset of $[0, T]$.
- A necessary and sufficient condition for the system to start in regime 2 is

$$\frac{1-e^{-a_{12}T}}{a_{12}} + \frac{1-e^{-a_{23}T}}{a_{23}} + \frac{e^{-a_{12}T} - e^{-a_{23}T}}{a_{12} - a_{23}} < \frac{1-e^{-a_{13}T}}{a_{13}}$$

- A necessary and sufficient condition for the system to start in regime 3 is

$$\frac{1-e^{-a_{21}T}}{a_{21}} + \frac{1-e^{-a_{13}T}}{a_{13}} + \frac{e^{-a_{21}T} - e^{-a_{13}T}}{a_{21} - a_{13}} < \frac{1-e^{-a_{23}T}}{a_{23}}$$

4.3.2 Optimal Routing Algorithm: Implementation

From the above discussions, it follows that the optimal routing algorithm can be implemented on a network in the following manner

Step 1: Estimate the link parameters a_{12}, a_{13}, a_{21}, and a_{23} of the network

Step 2: Check if they satisfy the following inequality

$$\frac{1-e^{-a_{12}T}}{a_{12}} + \frac{1-e^{-a_{23}T}}{a_{23}} + \frac{e^{-a_{12}T} - e^{-a_{23}T}}{a_{12} - a_{23}} < \frac{1-e^{-a_{13}T}}{a_{13}}$$

If the above is satisfied, then compute the switching instant t_s from the equation

$$\frac{1-e^{a_{13}(t_s-T)}}{a_{13}} = \frac{1-e^{a_{12}(t_s-T)}}{a_{12}} + \frac{1-e^{a_{23}(t_s-T)}}{a_{23}} + \frac{e^{a_{12}(t_s-T)} - e^{a_{23}(t_s-T)}}{a_{12} - a_{23}}$$

The routing strategy is specified as .

$\alpha_{12}(t) \equiv 1 \;\; \alpha_{21}(t) \equiv 0$, for $t \in [0, t_s)$,

$$\alpha_{12}(t) \equiv 0, \alpha_{21}(t) \equiv 0, \text{ for } t \in [t_s, T]$$

On the contrary, if the link parameters satisfy the following inequality

$$\frac{1-e^{-a_{21}T}}{a_{21}} + \frac{1-e^{-a_{13}T}}{a_{13}} + \frac{e^{-a_{21}T}-e^{-a_{13}T}}{a_{21}-a_{13}} < \frac{1-e^{-a_{23}T}}{a_{23}},$$

then compute the switching instant t_s from the equation:

$$\frac{1-e^{a_{23}(t_s-T)}}{a_{23}} = \frac{1-e^{a_{21}(t_s-T)}}{a_{21}} + \frac{1-e^{a_{13}(t_s-T)}}{a_{13}} + \frac{e^{a_{21}(t_s-T)}-e^{a_{13}(t_s-T)}}{a_{21}-a_{13}}$$

The optimal routing strategy is specified as:

$$\alpha_{12}(t) \equiv 0, \alpha_{21}(t) \equiv 1, \text{ for } t \in [0, t_s),$$
$$\alpha_{12}(t) \equiv 0, \alpha_{21}(t) \equiv 0, \text{ for } t \in [t_s, T].$$

Step 3: If neither of the two conditions of step 2 are satisfied, then the optimal routing strategy is

$$\alpha_{12}(t) \equiv 0, \alpha_{21}(t) \equiv 0, \forall t \in [0, T]$$

There are some interesting features of the above results regarding the network operation under the optimal routing strategy If the network operation is confined to the linear mode for the entire duration, then the routing strategy depends only on the topological parameters of the network. Such a case where all the links operate in the linear mode often arises in practical situations, if the traffic load on the network is low, and/or if the channel capacities of all the links are very high A topologically determined routing strategy can be easily implemented on an on line basis.

Another interesting feature of the optimal routing strategy under the linear mode of operation of all the links is that during no interval of the network operation, packets are routed onto both the links (1,2) and (2,1) simultaneously This property of the optimal routing strategy, which we term as the **loop-free property** will be shown to hold good even when one of the direct links of network is of finite channel capacity We prove this result in the Section 4.4.

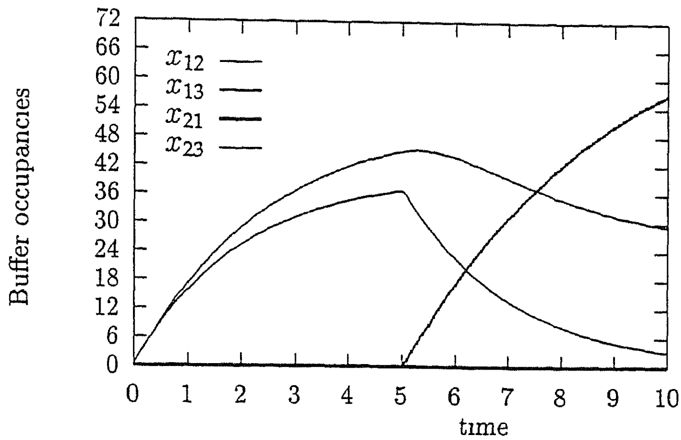

Figure 4.2. State variables as functions of time for case 1.

4.3.3 Numerical Examples

The above results regarding the nature of the optimal routing strategy can be illustrated by means of the following three cases

Case 1 : Consider a network with the following link parameters

$$a_{12} = 0\ 5, a_{13} = 0\ 25, a_{21} = 0\ 2, a_{23} = 0\ 8.$$

The initial buffer occupancies are assumed to be zero. The input traffic $\lambda_1(t) = \lambda_2(t) = 20$ packets/unit time

It can be easily verified that the above link parameters satisfy the necessary and sufficient condition (given by Theorem 4 3 4) for the network operation to start in regime 2. The system of differential equations in the state and costate variables (which in this case forms a (decoupled) two point boundary value problem) is solved numerically. Figure 4 2 shows the buffer occupancies as functions of time for this case Figure 4 3 is the costate variables corresponding to this case The network operates in regime 2 in the interval $[0, 5]$ and subsequently in regime 4 in the interval $(5, 10]$. Figure 4 4 shows the variation in the performance index J as the switching instant is varied over the interval of network operation.

Case 2 : Consider a network with the following link parameters:

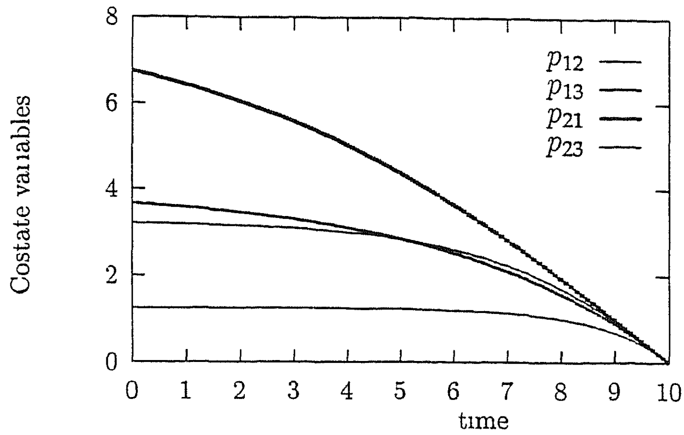

Figure 4 3: Costate variables as functions of time for case 1

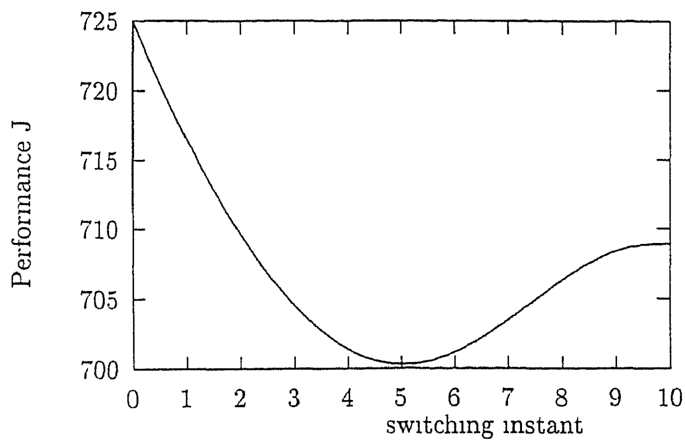

Figure 4.4: Performance variation with the switching instant for case 1

$$a_{12} = 0\ 3, a_{13} = 0\ 6, a_{21} = 0\ 9, a_{23} = 0\ 2.$$

The initial buffer occupancies are assumed to be zero The input traffic $\lambda_1(t) = \lambda_2(t) = 20$ packets/unit time.

The network operates in regime 3 in $[0, 8\ 1)$ and in regime 4 during the interval $[8\ 1, 10]$. The above choice of the link parameters satisfy the necessary and sufficient condition for the network operation to start in regime 3 Figure 4.5 shows the buffer

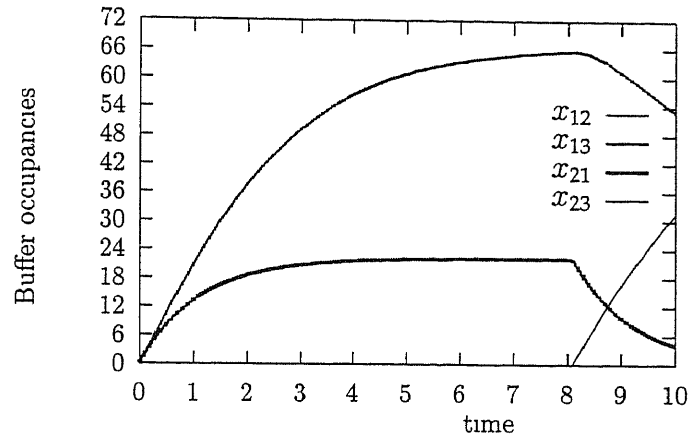

Figure 4.5 State variables as functions of time for case 2.

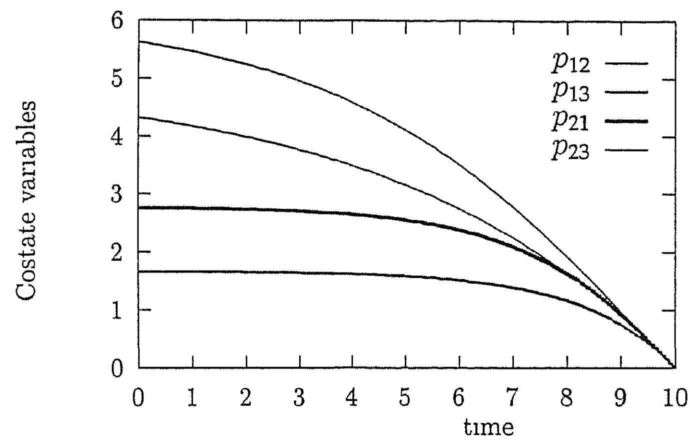

Figure 4 6 Costate variables as functions of time for case 2.

occupancies as functions of time for this case. Figure 4 6 shows the costate variables as functions of time. The variation in the performance index J as the switching instant (from regime 3 to regime 4) is varied over the interval of network operation $[0, 10]$, is shown in Figure 4 7

Case 3 : Consider a network with the following link parameters

$$a_{12} = 0\ 4, a_{13} = 0\ 5, a_{21} = 0\ 2, a_{23} = 0\ 3.$$

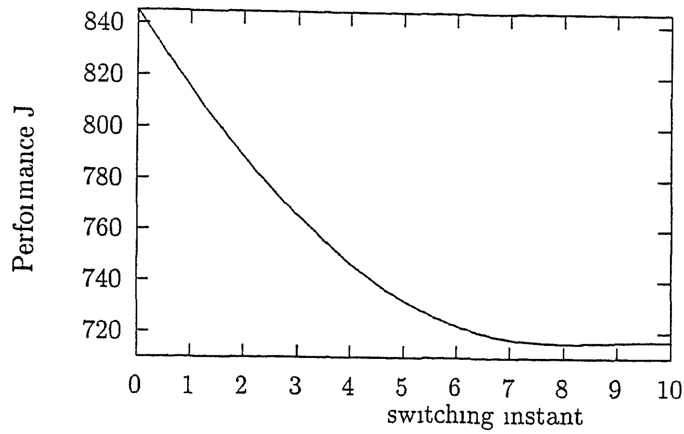

Figure 4 7 Performance variation with the switching instant for case 2

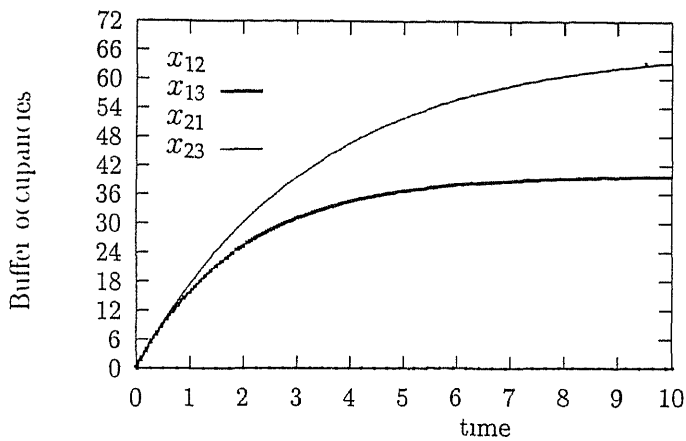

Figure 4.8 State variables as functions of time for case 3

The initial buffer occupancies are zero and the input traffic $\lambda_1(t) = \lambda_2(t = 20$ packets/unit time.

Figure 4.8 shows the buffer occupancies and Figure 4 9 shows the costate variables for this case The choice of the link parameters result in a regime 4 operation for the entire interval $[0, 10]$

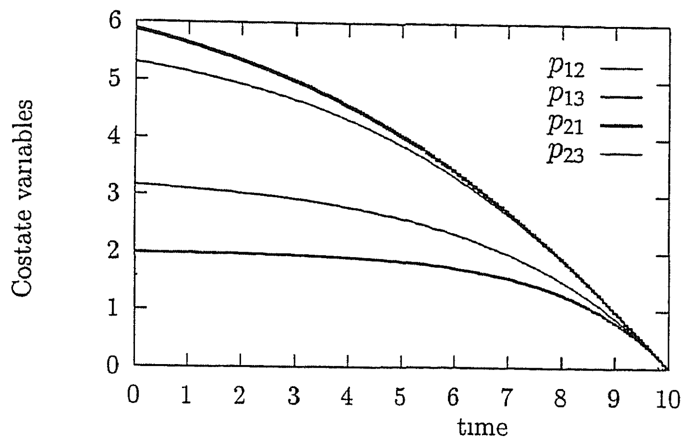

Figure 4.9: Costate variables as functions of time for case 3

4.4 Networks with a Direct Link of Finite Channel Capacity

In this section, we examine the properties of the optimal routing strategy for the three-node network in which one of the direct links has finite channel capacity. Using an argument similar to the one used in the previous chapter for the two-node case, we show that the routing strategy at a source node [the one which has a direct link is of finite channel capacity] need not be *bang-bang*. There could be intervals in the time span of network operation during which this routing variable [either $\alpha_{13}(t)$ or $\alpha_{23}(t)$] takes non-zero, non-unity values. We prove that over such intervals the value of this variable depends on the channel capacity of the associated link and on the total input traffic to that node. We also prove some interesting results pertaining to the possible modes in which the network operation can end, and provide numerical examples to substantiate these results. Furthermore, we establish the **loop-free** property, as in the linear case of the previous section. These results are stated as theorems in the subsections that follow.

In the analysis that follows we assume that link (1,3) has finite channel capacity. All other links of the network are assumed to be of infinite channel capacity.

The dynamic evolution of the state of the network is then given by the following set of differential equations:

$$x_{12} = -a_{12}x_{12}(t) + \alpha_{12}(t)[\lambda_1(t) + a_{21}x_{21}(t)] \tag{4.23}$$

$$\dot{x}_{13} = -f_{13}(x_{13}) + \alpha_{13}(t)[\lambda_1(t) + a_{21}x_{21}(t)] \tag{4.24}$$

$$x_{21} = -a_{21}x_{21}(t) + \alpha_{21}(t)[\lambda_2(t) + a_{12}x_{12}(t)] \tag{4.25}$$

$$x_{23} = -a_{23}x_{23}(t) + \alpha_{23}(t)[\lambda_2(t) + a_{12}x_{12}(t)] \tag{4.26}$$

where the function $f_{13}(x_{13})$ is defined as

$$f_{13}(x_{13}) = \begin{cases} a_{13}x_{13} & \text{if } x_{13} < C_{13}/a_{13} \\ C_{13} & \text{otherwise} \end{cases} \tag{4.27}$$

The routing variables $\alpha_{12}(t)$, $\alpha_{13}(t)$, $\alpha_{21}(t)$ and $\alpha_{23}(t)$ have the same interpretation as in the previous section , and the normalizing conditions which they must satisfy remain unchanged from those given in the Section 4 2

The performance index J, that is to be minimised is the total buffer occupancy time defined (as in the earlier section) below:

$$J = \int_0^T (x_{12}(t) + x_{13}(t) + x_{21}(t) + x_{23}(t))dt$$

Thus the problem of finding the optimal routing strategy is formulated as the optimal control problem of finding the control variables $\alpha_{12}(t)$, $\alpha_{13}(t)$, $\alpha_{21}(t)$ and $\alpha_{23}(t)$ for the system given by the Equations (4 23),(4.24),(4.25) and (4.26), which minimise the performance index J

We investigate the above optimal control problem in the framework of Pontryagin's Maximum principle, as done in the previous Section 4 1. Since the function $f_{13}(x_{13})$ has a discontinuity at x_{13s} we consider a class of control problems in which the function $f_{13}(x_{13})$ is replaced by $f^r_{13}(x^r_{13})$ defined as below

$$f^r_{13}(x_{13}) = \begin{cases} a_{13}x_{13} & \text{if } x_{13} \le x_{13l} = \frac{C_{13}}{a_{13}} + \frac{r(1-\sqrt{1+a_{13}2}}{a_{13}\sqrt{1+a_{13}^2}} \\ \sqrt{r^2 - (x_{13} - x_{13c})^2} + C_{13} - r, & \text{if } x_{13l} < x_{13}(t) < x_{13c} \\ C_{13} & \text{if } x_{13} \ge x_{13c} \end{cases}$$

where $x_{13c} = \frac{C_{13}}{a_{13}} + \frac{r(\sqrt{(1+a_{13}2)}-1)}{a_{13}}$ and $r > 0$

Comments:

As in the case of Chapter 3 the function $f_{13}^r(x_{13})$ is obtained by drawing an arc of radius r, tangential to the lines $f_{13}(x_{13}) = a_{13}x_{13}$ and $f_{13}(x_{13}) = C_{13}$

And as $r \to 0$,

(i) $x_{13l}, x_{13c} \to x_{13s}$

(ii) $f_{13}^r(x_{13}) \to f_{13}(x_{13}), \forall x_{13} \geq 0$.

The function $\frac{df_{13}^r(x_{13})}{dx_{13}}$ is continuous in x_{13} ($\forall x_{13} \geq 0$) and is given by

$$\frac{df_{13}^r(x_{13})}{dx_{13}} = \begin{cases} a_{13} & \text{if } x_{13} \leq x_{13l} \\ \frac{x_{13c}-x_{13}}{\sqrt{r^2-(x_{13c}-x_{13})^2}} & \text{if } x_{13l} \leq x_{13c} \leq x_{13c} \\ 0 & \text{if } x_{13} > x_{13c} \end{cases}$$

The Hamiltonian for the above class of problems is given as

$H^r(\underline{x}^r, \underline{p}^r, \underline{\alpha}^r) = x_{13}^r(t) + x_{12}^r(t) + x_{21}^r(t) + x_{23}^r(t) + p_{13}^r \dot{x}_{13}^r + p_{12}^r \dot{x}_{12}^r + p_{21}^r \dot{x}_{21}^r + p_{23}^r \dot{x}_{23}^r$

Substituting the dynamics given by Equations (4 23)-(4 27), we get

$$\begin{aligned} H^r(\underline{x}^r, \underline{p}^r, \underline{\alpha}^r) &= [x_{13}^r(t) + x_{12}^r(t) + x_{21}^r(t) + x_{23}^r(t)] \\ &+ (p_{12}^r - p_{13}^r)\alpha_{12}^r(t)[\lambda_1(t) + a_{21}x_{21}^r(t)] \\ &+ (p_{21}^r - p_{23}^r)\alpha_{21}^r(t)(\lambda_2(t) + a_{12}x_{12}^r(t)) \\ &- a_{12}p_{12}^r x_{12}^r + p_{13}^r(\lambda_1(t) + a_{21}x_{21}^r) \\ &+ p_{23}^r(\lambda_2(t) + a_{12}x_{12}) - f_{13}^r(x_{13}^r)p_{13}^r \\ &- a_{23}p_{23}^r x_{23}^r - a_{21}x_{21}^r p_{21}^r \end{aligned}$$

Minimising the Hamiltonian w r t. $\alpha_{12}^r(t)$ and $\alpha_{21}^r(t)$ leads to the following conditions

(1) If $p_{12}^r(t) > p_{13}^r(t)$, then $\alpha_{12}^r(t) = 0$ and $\alpha_{13}^r(t) = 1$

(2) If $p_{12}^r(t) < p_{13}^r(t)$, then $\alpha_{12}^r(t) = 1$ and $\alpha_{13}^r(t) = 0$

(3) If $p^r_{21}(t) > p^r_{23}(t)$ then $\alpha^r_{21}(t) = 0$ and $\alpha^r_{23}(t) = 1$

(4) If $p^r_{21}(t) < p^r_{23}(t)$, then $\alpha^r_{21}(t) = 1$, and $\alpha^r_{23}(t) = 0$

If $\forall t \in I$, $p^r_{12}(t) = p^r_{13}(t)$, then $\alpha^r_{12}(t) = 1$ and $\alpha^r_{13}(t)$ are not specified by conditions (1) and (2).

If $\forall t \in I$, $p^r_{21}(t) = p^r_{23}(t)$, then $\alpha^r_{21}(t)$ is not specified by conditions (3) and (4).

According to the Maximum principle, the optimal state and costate variables satisfy the following necessary conditions:

(i) $\dot{p}^r_{jk}(t) = -\frac{\partial H^r}{\partial x^r_{jk}}$ with the boundary conditions $p^r_{jk}(T) = 0$.

(ii) $\dot{x}^r_{jk} = \frac{\partial H^r}{\partial p^r_{jk}}$

From (i) above, the differential equations for the costate variables $p^r_{jk}(t)$ are given as

$$\dot{p}^r_{12} = -1 + a_{12}p^r_{12} - a_{12}\alpha^r_{21}p^r_{21} - a_{12}\alpha^r_{23}p^r_{23} \tag{4.28}$$

$$\dot{p}^r_{13} = -1 + \frac{df^r_{13}}{dx^r_{13}}p^r_{13} \tag{4.29}$$

$$\dot{p}^r_{21} = -1 + a_{21}p^r_{21} - a_{21}\alpha^r_{12}p^r_{12} - a_{21}\alpha^r_{13}p^r_{13} \tag{4.30}$$

$$\dot{p}^r_{23} = -1 + a_{23}p^r_{23} \tag{4.31}$$

along with the boundary conditions $p^r_{12}(T) = p^r_{13}(T) = p^r_{21}(T) = p^r_{23}(T) = 0$

Equation (4.31) can be solved independently to obtain

$$p^r_{23}(t) = \frac{1 - e^{a_{23}(t-T)}}{a_{23}}$$

From the differential Equations (4.28)-(4.31) in the costate variables, certain interesting observations which are stated below as lemmas, can be made regarding the nature of the optimal routing strategy

Lemma 4.4.1 *There exists no interval $I = [t_0, t_1]$, $0 \leq t_0 < t_1 < T$ such that $\forall t \in I$, $p^r_{12}(t) = p^r_{13}(t)$ and $p^r_{21}(t) = p^r_{23}(t)$.*

Proof : (By contradiction)

Let $I = [t_0, t_1]$ be an interval during which $p^r_{12}(t) \equiv p^r_{13}(t)$ and $p^r_{21}(t) \equiv p^r_{23}(t)$.
Then $\dot{p}^r_{12} \equiv \dot{p}^r_{13}$ and $\dot{p}^r_{21} \equiv \dot{p}^r_{23}$, $\forall t \in I$
This implies that $\forall t \in I$

$$\begin{aligned} a_{12}(p^r_{12} - p^r_{21}) &= \frac{df^r_{13}}{dx^r_{13}} p^r_{13} \\ a_{21}(p^r_{21} - p^r_{12}) &= a_{23} p^r_{23} \\ \text{Hence} \quad \frac{df^r_{13}}{dx^r_{13}} &= -\left(\frac{a_{12}}{a_{21}} a_{23}\right) \frac{p^r_{23}}{p^r_{13}} \quad \ldots\ldots\,.. (A). \end{aligned}$$

It can be argued as follows, that the right hand side of the Equation (A) is negative. The function $p^r_{23}(t)$ which is given as

$$p^r_{23}(t) = \frac{1 - e^{a_{23}(t-T)}}{a_{23}}$$

is positive in $[0, T)$ The function $p^r_{13}(t)$ is also positive in $[0, T)$ since

$$p^r_{13}(t) = [\int_t^T exp^{-\int_0^\tau \frac{df^r_{13}(x^r_{13})}{dx^r_{13}} d\gamma} d\tau][exp^{\int_0^t \frac{df^r_{13}(x^r_{13}(\tau))}{dx^r_{13}} d\tau}]$$

is a product of two terms which are both positive in $[0\ T)$ The link parameters a_{12}, a_{21} and a_{23} are all positive constants
The left hand side of the Equation (A) is non-negative since $\frac{df^r_{13}}{dx^r_{13}} \geq 0$, $\forall t \in [0, T]$
Hence Equation (A) can not hold true in any interval I which is a subset of $[0, T)$
Thus the assumption of the existence of an interval in which $p^r_{12}(t) \equiv p^r_{13}(t)$ and $p^r_{21}(t) \equiv p^r_{23}(t)$ leads to a contradiction

□

Lemma 4.4.2 *There exists no interval $I = [t_0, t_1]$, $0 \leq t_0 < t_1 \leq T$, such that $\forall t \in I$, $p^r_{21}(t) = p^r_{23}(t)$ and $p^r_{12}(t) < p^r_{13}(t)$*

Proof : Let us assume that during some interval $I = [t_0, t_1]$, $p^r_{21}(t) \equiv p^r_{23}(t)$ and $p^r_{12}(t) < p^r_{13}(t)$ The differential equations, for $p^r_{12}(t)$and $p^r_{21}(t)$ in the interval I are given as:

$$\dot{p}^r_{12} = -1 + a_{12}(p^r_{12} - p^r_{21})$$

$$\text{and} \quad \dot{p}^r_{21} = -1 + a_{21}(p^r_{21} - p^r_{12}).$$

Since $p^r_{21} \equiv p^r_{23}$, it follows that $\dot{p}^r_{21} \equiv \dot{p}^r_{23}$ during I

$$\text{Therefore} \quad a_{21}(p^r_{21} - p^r_{12}) = a_{23}p^r_{23}$$
$$\text{i e} \quad p^r_{12} = \frac{(a_{21} - a_{23})}{a_{21}} p^r_{23}$$

Consider the case where $a_{21} = a_{23}$.

$$\begin{aligned} a_{21} = a_{23} \;\Rightarrow\; & p^r_{12}(t) \equiv 0, \forall t \in I. \\ \Rightarrow \dot{p}^r_{12} \;=\; & -1 + a_{12}p^r_{12} - a_{12}p^r_{21} \\ =\; & -(1 + a_{12}p^r_{21}) < 0 \quad \forall t \in I \; .....(\mathrm{B}) \end{aligned}$$

But inequality (B) can not hold true since, if $p^r_{12}(t)$ is identically equal to zero during the entire interval I then $\dot{p}^r_{12}$ also has to be zero during that interval

Consider the case where $a_{21} \neq a_{23}$.

$$\begin{aligned} p^r_{12}(t) &= \frac{(a_{21} - a_{23})}{a_{21}} p^r_{23}(t) \\ &= \frac{(a_{21} - a_{23})}{a_{21}} p^r_{21}(t) \\ p^r_{12}(t) &= -\frac{(a_{21} - a_{23})}{a_{21}} e^{a_{23}(t-T)} \\ 1 + p^r_{12}(t) &= 1 - \frac{(a_{21} - a_{23})}{a_{21}} e^{a_{23}(t-T)} \end{aligned}$$

From the differential equation for the costate variable $\dot{p}^r_{12}(t)$, we know that

$$\begin{aligned} 1 + p^r_{12}(t) &= a_{12}(p^r_{12}(t) - p^r_{21}(t)) \\ \text{Hence} \quad 1 - ((a_{21} - a_{23})/a_{21})e^{a_{23}(t-T)} &= a_{12}(p^r_{12}(t) - p^r_{21}(t)) \\ &= a_{12}\left(\frac{(a_{21} - a_{23})}{a_{21}} - 1\right)\frac{(1 - e^{a_{23}(t-T)})}{a_{23}} \end{aligned}$$

After rearranging the terms in the above equation, we get the following:

$$\frac{(1/a_{12} + 1/a_{21})}{((a_{21} - a_{23})/a_{12}a_{21}) + 1/a_{21}} = e^{a_{23}(t-T)}, \quad \forall t \in I \quad \ldots(\mathrm{C})$$

Note that the left side of Equation (C) is a constant while the right side is a monotonically increasing function of time and therefore it can not be satisfied. Thus

from the assumption of the existence of an interval I during which $p^r_{21}(t) \equiv p^r_{23}(t)$ and $p^r_{12}(t) < p^r_{13}(t)$, we arrive at a contradiction in either of the two possible cases $[a_{21} = a_{23}$ or $a_{21} \neq a_{23}]$. Hence the lemma.

□

Lemma 4.4.3 *There exists no interval $I_0 = [t_0, t_1]$, $0 \leq t_0 < t_1 \leq T$, such that $\forall t \in I_0$, $p^r_{12}(t) = p^r_{13}(t)$ and $p^r_{21}(t) < p^r_{23}(t)$.*

Proof : Let $I_0 = [t_0, t_1]$ be an interval during which $p^r_{12}(t) \equiv p^r_{13}(t)$ and $p^r_{21}(t) < p^r_{23}(t)$. The routing variable $\alpha^r_{23}(t)$ is identically equal to zero in I_0 and therefore the dynamics of $p^r_{12}(t)$ and $p^r_{21}(t)$ in I_0 are given as:

$$\dot{p}^r_{12} = -1 + a_{12}[p^r_{12} - p^r_{21}]$$
$$\dot{p}^r_{21} = -1 + a_{21}[p^r_{21} - p^r_{12}]$$

Since $p^r_{12} \equiv p^r_{13}$, $\dot{p}^r_{12} \equiv \dot{p}^r_{13}$ in I_0.

We know that $\dot{p}^r_{13} = -1 + \frac{df^r_{13}}{dx^r_{13}} p^r_{13} \geq -1$ (since $\frac{df^r_{13}}{dx^r_{13}} \geq 0$ and $p^r_{13} \geq 0$).

Thus $\dot{p}^r_{12} \geq -1$, which in turn implies that $p^r_{12}(t) \geq p^r_{21}(t)$, and $\dot{p}^r_{21} \leq -1$, $\forall t \in I_0$.

Thus the following statements hold true :

For all t in the interval I_0,

(a) $p^r_{12}(t) \geq p^r_{21}(t)$

(b) $p^r_{13}(t) \geq p^r_{21}(t)$

(c) $\dot{p}^r_{21}(t) \leq -1$

(d) $\dot{p}^r_{23} - \dot{p}^r_{21} \geq a_{23} p^r_{23} > 0$.

From (d) it can be concluded that $(p^r_{23}(t) - p^r_{21}(t))$ is a monotonically increasing function (and is positive as per the assumption that during I_0, $p^r_{23}(t) > p^r_{21}(t)$) in I_0. Therefore

1. I_0 can not be a terminating regime since $p^r_{23}(T) > p^r_{21}(T)$ violates the transversality condition.

2. I_0 is followed[1] by an interval (howsoever small) in which $p^r_{21}(t) < p^r_{23}(t)$ and

[1]Implicit in this inference is the fact that the costate variables are continuous functions of time in $[0, T]$

$p_{12}^r(t) \neq p_{13}^r(t)$.

Let $I_1 = [t_1, t_2]$ where $t_1 < t_2 < T$ (for some t_2), denote this interval that follows I_0 during which $p_{21}^r(t)$ continues to be less than $p_{23}^r(t)$. We shall examine the nature of the costate variables in I_1.

Note: A transition from an interval I_0 in which $p_{21}^r(t) < p_{23}^r(t)$ and $p_{12}^r(t) \equiv p_{13}^r(t)$ to an interval in which $p_{21}^r(t) < p_{23}^r(t)$ and $p_{12}^r(t) \neq p_{13}^r(t)$ can arise only from two possibilities : case 1, in which $p_{12}^r(t) < p_{13}^r(t)$, and case 2, in which $p_{12}^r(t) > p_{13}^r(t)$.

Case 1:

If $p_{12}^r(t) < p_{13}^r(t)$ in I_1, then the dynamics of $p_{12}^r(t)$ and $p_{21}^r(t)$ in I_1 are given as

$$\begin{aligned} \dot{p}_{12}^r &= -1 + a_{12}(p_{12}^r - p_{21}^r) \\ \dot{p}_{21}^r &= -1 + a_{21}(p_{21}^r - p_{12}^r) \end{aligned}$$

with the initial conditions $p_{12}^r(t_1) \geq p_{21}^r(t_1)$ [this follows from (a)].

$$\begin{aligned} \text{Since } \dot{p}_{12}^r - \dot{p}_{21}^r &= (a_{12} + a_{21})(p_{12}^r - p_{21}^r) \\ p_{12}^r(t) - p_{21}^r(t) &= [p_{12}^r(t_1) - p_{21}^r(t_1)]e^{(a_{12}+a_{21})(t-t_1)} \geq 0, \quad \forall t \in I_1. \\ \text{and consequently } \dot{p}_{21}^r(t) &\leq -1, \quad \forall t \in I_1. \end{aligned}$$

Since $\qquad \dot{p}_{23}^r = -1 + a_{23}p_{23}^r > -1, \qquad \dot{p}_{23}^r - \dot{p}_{21}^r > 0 \; \forall t \in I_1.$

Hence $(p_{23}^r(t) - p_{21}^r(t))$ continues to be a **monotonically increasing function in I_1.**

$$\begin{aligned} \text{Furthermore } p_{21}^r(t) &\leq p_{12}^r(t) < p_{13}^r(t). \\ \text{i.e. } p_{21}^r(t) &\leq \min[p_{12}^r(t), p_{13}^r(t)] \quad \forall t \in I_1. \end{aligned}$$

Case 2:

If $p_{12}^r(t) > p_{13}^r(t)$ in I_1, then the dynamics of $p_{12}^r(t)$ and $p_{21}^r(t)$ in I_1 are given as

$$\begin{aligned} \dot{p}_{12}^r &= -1 + a_{12}[p_{12}^r - p_{21}^r] \\ \dot{p}_{21}^r &= -1 + a_{21}[p_{21}^r - p_{13}^r] \\ \text{Since, } \dot{p}_{13}^r &= -1 + \frac{df_{13}^r}{dx_{13}^r} p_{13}^r \\ \dot{p}_{21}^r - \dot{p}_{13}^r &= a_{21}[p_{21}^r - p_{13}^r] - \frac{df_{13}^r}{dx_{13}^r} p_{13}^r \leq a_{21}[p_{21}^r - p_{13}^r]. \end{aligned}$$

At $t = t_1$ the initial conditions of $p^r_{21}(t_1)$ and $p^r_{13}(t_1)$ are such that $p^r_{21}(t_1) \leq p^r_{13}(t_1)$ [from (b)]. It can then be argued that $p^r_{21}(t) - p^r_{13}(t) \leq 0,\ \forall t \in I_1$. Therefore $\dot{p}^r_{21} = -1 + a_{21}[p^r_{21} - p^r_{13}] \leq -1$ and $(p^r_{23}(t) - p^r_{21}(t))$ will continue to be **monotonically increasing** $\forall t \in I_1$.

$$\begin{aligned} \text{Furthermore,} \quad p^r_{21}(t) &\leq p^r_{13}(t) < p^r_{12}(t) \\ \text{i.e.} \quad p^r_{21}(t) &\leq \min[p^r_{13}(t), p^r_{12}(t)]. \end{aligned}$$

Thus in both the cases 1 and 2, all the four statements (a), (b), (c) and (d) (regarding the nature of the costate variables in the interval I_0) hold true in I_1 also.

The above arguments which were used to prove that I_0 is followed by an interval I_1 during which the properties stated in (a), (b), (c) and (d) continue to hold true, can be extended to show that I_1 in turn is also followed by such an interval. Extending the above reasonings to regimes that follows I_1 (in succession), it can be concluded that $p^r_{23}(t) > p^r_{21}(t),\ \forall t \in [\ t_0, T]$. Since this results in a condition $p^r_{23}(T) > p^r_{21}(T)$ which violates the transversality condition, the initial assumption of the existence of an interval I_0 with the stated properties is incorrect. Hence the lemma.

□

4.4.1 Routing Strategy at node 1

From the three Lemmas 4.3.1, 4.3.2 and 4.3.3 it can be concluded that during any interval in which the costate variables $p^r_{12}(t)$ and $p^r_{13}(t)$ are identically equal, the costate variable $p^r_{21}(t)$ has to be greater than $p^r_{23}(t)$ and consequently the optimal routing variable $\alpha^r_{21}(t)$ has to be zero during that interval. This implies that at the source node 2, all the incoming traffic is routed onto the direct link (2,3) during such an interval. We now establish that at the source node 1, (for the case of r tending to zero) a fraction (of the incoming traffic) equal to the ratio of the channel capacity of link (1,3) to the total traffic arriving at node 1 is routed onto the direct link (1,3) during such an interval.

Theorem 4.4.1 *For the control problem in which r tends to zero, the only possible values* $\alpha^r_{13}(t)$ *can take are 0, 1 and* $\frac{C_{13}}{\lambda_1(t)+a_{21}x^r_{21}(t)}$

Proof : It has already been argued that if during an interval $p^r_{12}(t)$ is greater than $p^r_{13}(t)$, then $\alpha^r_{13}(t)$ equals unity and if $p^r_{12}(t)$ is less than $p^r_{13}(t)$, then $\alpha^r_{13}(t)$ equals zero in that interval Let $p^r_{12}(t) = p^r_{13}(t), \forall t \in I \subseteq [0,T]$. Then, by Lemmas 4 4 1 and 4.4.3, $p^r_{21}(t) > p^r_{23}(t), \forall t \in I$.

$$\text{Therefore} \quad \dot{p}^r_{12}(t) = -1 + a_{12}(p^r_{12}(t) - p^r_{23}(t))$$

Since $p^r_{12}(t) \equiv p^r_{13}(t)$ in I, $\dot{p}^r_{12} \equiv \dot{p}^r_{13}, \forall t \in I$

$$\text{Hence} \quad \frac{df^r_{13}}{dx^r_{13}} = \frac{a_{12}(p^r_{12} - p^r_{23})}{\dot{p}^r_{12}} \quad \text{during } I$$

$\frac{df^r_{13}}{dx^r_{13}}$ is thus a function of t, and not a constant during I

Comments: This follows from the nature of the solution for $p^r_{12}(t)$ of the differential equation in $p^r_{12}(t)$

Therefore during I, $x_{13l} \leq x^r_{13}(t) \leq x_{13c}$ In the limiting case of r tending to zero, both x_{13l} and x_{13c} tend to x_{13s} Therefore $x^r_{13}(t)$ has to be identically equal to x_{13s}, during I and $\dot{x}^r_{13}$ is zero for the entire interval Consequently $\alpha^r_{13}(t)$ equals $\frac{C_{13}}{\lambda_1(t)+a_{21}x^r_{21}(t)}$ over an interval during which $p^r_{12}(t)$ is identically equal to $p^r_{21}(t)$ for the case where r tends to zero.

□

In the Section 4 5 4 we shall provide numerical examples wherein the routing variable $\alpha_{13}(t)$ takes the value $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$ during sub-intervals of network operation We now prove the loop-free property of the routing strategy.

4.4.2 Loop-free Property

We have established in Section 4 3 that both the links (1,2) and (2,1) are never simultaneously used to carry the incoming traffic at node 1 and node 2 respectively

when all the links of the network are of infinite channel capacity. Making use of the above three lemmas, we can prove that such is the case even when a direct link is of finite channel capacity.

Theorem 4.4.2 *There exists no interval $I \subseteq [0,T]$ during which both $\alpha^r_{12}(t)$ and $\alpha^r_{21}(t)$ are non-zero*

Proof : By Lemmas 4.4.1 and 4.4.3, if $p^r_{12}(t)$ and $p^r_{13}(t)$ are identically equal in an interval I, then $p^r_{21}(t)$ is greater than $p^r_{23}(t)$ during I and consequently the routing variable $\alpha^r_{21}(t)$ is identically equal to zero in I. We shall now establish (via a contradiction) that if $p^r_{12}(t)$ is less that $p^r_{13}(t)$ in an interval I, then $p^r_{21}(t)$ has to be greater than $p^r_{23}(t)$ during such an interval.

Let $I = [t_0, t_1]$ where $0 \leq t_0 < t_1 \leq T$ be an interval such that $p^r_{12}(t) < p^r_{13}(t)$ and $p^r_{21}(t) \leq p^r_{23}(t)$ for the entire interval

Then during I, the dynamics of $p^r_{12}(t)$ and $p^r_{21}(t)$ are given as

$$\begin{aligned}
\dot{p}^r_{12} &= -1 + a_{12}(p^r_{12} - p^r_{21}) \\
\dot{p}^r_{21} &= -1 + a_{21}(p^r_{21} - p^r_{12}) \\
\text{Therefore} \quad (p^r_{12}(t) - p^r_{21}(t)) &= (p^r_{12}(t_0) - p^r_{21}(t_0))e^{(a_{12}+a_{21})(t-t_0)}
\end{aligned}$$

Consider the case where $p^r_{12}(t_0) \geq p^r_{21}(t_0)$

Then $p^r_{12}(t) \geq p^r_{21}(t), \forall t \in I$ and consequently $\dot{p}^r_{21}(t) \leq -1, \forall t \in I$

Since $\dot{p}^r_{23}(t) > -1$, $\dot{p}^r_{23}(t) - \dot{p}^r_{21}(t) > 0 \; \forall t \in I$ and $(p^r_{23}(t) - p^r_{21}(t))$ is a monotonically increasing function of time. Besides $p^r_{21}(t) \leq p^r_{12}(t) < p^r_{13}(t)$.

i.e. $\quad p^r_{21}(t) \leq min(p^r_{12}(t), p^r_{13}(t))$

We have already argued in the proof of Lemma 4.4.3 (see case 1) that if the costate variables $p^r_{12}(t)$, $p^r_{13}(t)$, $p^r_{21}(t)$ and $p^r_{23}(t)$ in an interval I are such that

(i) $p^r_{12}(t) < p^r_{13}(t), \forall t \in I$,

(ii) $(p^r_{23}(t) - p^r_{21}(t))$ is a positive and monotonically increasig function in I,

(iii) $p^r_{21}(t) \leq \min \{p^r_{12}(t), p^r_{13}(t)\}, \forall t \in I$

then the relationship $p_{23}^r(t) > p_{21}^r(t)$ continues to hold true for all t upto T. This results in the condition $p_{23}^r(T) > p_{21}^r(T)$ which violates the transversality condition $p_{23}^r(T) = p_{21}^r(T) = 0$.

Consider the case wherein $p_{21}^r(t_0) > p_{12}^r(t_0)$

Then $p_{21}^r(t) > p_{12}^r(t), \forall t \in I$.

Therefore $p_{12}^r(t) < -1, \forall t \in I$

Since $p_{13}^r(t) \geq -1$, the function $(p_{13}^r(t) - p_{12}^r(t))$ which is positive (as per the assumption) is monotonically increasing in I

Furthermore, $p_{12}^r(t) < \min\{p_{21}^r(t), p_{23}^r(t)\}$

It can be argued from the above two conditions (we do not repeat the arguments) that they result in the condition that $p_{13}^r(t) > p_{12}^r(t), \forall t \in [t_0, T]$ and this in turn results in the violation of the transversality condition $p_{13}^r(T) = p_{12}^r(T) = 0$

Thus we arrive at a contradiction in both the two possible cases ($p_{12}^r(t_0) \geq p_{13}^r(t_0)$ or $p_{12}^r(t_0) < p_{13}^r(t_0)$), from the initial assumption of the existence of an interval during which $p_{12}^r(t)$ is less than $p_{13}^r(t)$ and $p_{21}^r(t)$ is less than or equal to $p_{23}^r(t)$. Therefore there exists no such interval in the duration of the network operation.

Thus if $p_{12}^r(t) \leq p_{13}^r(t)$ over any interval (i.e. $\alpha_{12}^r(t)$ is non-zero), then $\alpha_{21}^r(t)$ has to be zero during that interval.

□

4.4.3 Modes of Operation of the Network

We investigate the possible modes in which the network operation can end under the optimal routing strategy. The following definitions are used henceforth.

DEFINITIONS:

Definition 4.4.1 The network is said to be operating in the *saturation mode*(S) during an interval I if $\forall t \in I$, $x_{13}^r(t) \geq x_{13c}$, and such an interval I is termed as a *saturation regime*.

Definition 4.4.2 The network is said to be operating in the *linear mode* **(L)** during an interval I if $\forall t \in I, x_{13}^r(t) \leq x_{13l}$, and such an interval I is termed as a *linear regime*

Definition 4.4.3 The network is said to be operating in the *transition mode*(**T**) during an interval I if $\forall t \in I, x_{13l} < x_{13}^r(t) < x_{13c}$, and such an interval I is termed as a *transition regime.*

In the limiting case as r tends to zero, if $\forall t \in I, x_{13}(t) > x_{13s}$, then I is a *saturation regime*(**S**).
If $\forall t \in I$, $x_{13}(t) \equiv x_{13s}$, then I is a *transition regime*(**T**)
And if $\forall t \in I, x_{13}(t) < x_{13s}$, then I is a *linear regime* **(L).**

Definition 4.4.4 Link (1,2) (link (2,1)) is said to be operating in **Full (F)** during an interval I, if $\forall t \in I, \alpha_{12}^r(t)$ $(\alpha_{21}^r(t)) = 1$.

Definition 4.4.5 Link (1,2) (link (2,1)) is said to be operating in **Empty (E)** during I, if $\forall t \in I, \alpha_{12}^r(t)$ $(\alpha_{21}^r(t)) = 0$

Definition 4.4.6 Link (1,2) (link (2,1)) is said to be operating in **Partial (P)** during an interval I, if $\forall t \in I, 0 < \alpha_{12}^r(t)$ $(\alpha_{21}^r(t))$ < 1.

Denoting the mode of operation of the network by the triplet (eg FFS means $\alpha_{12}(t) = 1$, $\alpha_{21}(t) = 1$, and network in saturation mode) in which the first letter stands for the status of link (1,2), the second that of link (2,1) and the third, on the status of link (1,3), the following 27 modes are possible

FFS, FFT, FFL, FES, FET, FEL, FPS, FPT, FPL, EFS,EFT, EFL, EPS, EPT, EPL, EES,EET, EEL, PFS, PFT, PFL, PEL, PES, PET, PPS, PPT, PPL.

Of the 27 above, by the loop-free property (Theorem 4.4 2) the following are ruled out not only in a terminating interval, but also in any intermediate interval of

network operation.

FFS, FFT, FFL, PFS, PFT, PFL, FPS, FPT, FPL, PPS, PPT, PPL.

Since link (1,2) can be operational in partial mode (P) only if $p^r_{12}(t) \equiv p^r_{13}(t)$, we can rule out the modes PES and PEL in a terminal regime from the following considerations:

PES:

The dynamics of $p^r_{12}(t)$ and $p^r_{13}(t)$ are given as

$$\begin{aligned} \dot{p}^r_{12} &= -1 + a_{12}(p^r_{12} - p^r_{23}), \quad p^r_{12}(T) = 0 \\ \dot{p}^r_{13} &= -1, \quad p^r_{13}(T) = 0 \end{aligned}$$

The solutions to the above equations are

$$\begin{aligned} p^r_{12}(t) &= \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{(e^{a_{12}(t-T)} - e^{a_{23}(t-T)})}{(a_{12} - a_{23})} \\ p^r_{13}(t) &= T - t \end{aligned}$$

Since $p^r_{12}(t)$ and $p^r_{13}(t)$ as given by the above expressions can not be identically equal over any interval, link (1,2) can not be in partial mode Hence PES can not be the mode of operation in a terminal regime

PEL:

The dynamics of $p^r_{12}(t)$ and $p^r_{13}(t)$ are given as

$$\begin{aligned} \dot{p}^r_{12} &= -1 + a_{12}(p^r_{12} - p^r_{23}), \quad p^r_{12}(T) = 0. \\ \dot{p}^r_{13} &= -1 + a_{13}p^r_{13}, \quad p^r_{13}(T) = 0. \end{aligned}$$

The solution to the above equations are

$$\begin{aligned} p^r_{12}(t) &= \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{(e^{a_{12}(t-T)} - e^{a_{23}(t-T)})}{(a_{12} - a_{23})} \\ p^r_{13}(t) &= \frac{1 - e^{a_{13}(t-T)}}{a_{13}} \end{aligned}$$

Since $p^r_{12}(t)$ as given by the above expression can not be identically equal to $p^r_{13}(t)$ over an interval, link (1,2) can not be in partial mode Hence **PEL** is ruled out in a terminal regime.

The modes **FEL, EFS, EFL, EFL, EES** are also ruled out in a terminal regime

FEL:

$$\begin{aligned} p_{12}^r(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{(e^{a_{12}(t-T)} - e^{a_{23}(t-T)})}{(a_{12}-a_{23})}. \\ p_{13}^r(t) &= \frac{1-e^{a_{13}(t-T)}}{a_{13}} \end{aligned}$$

As proved in the proof for Theorem 4.3.3 (in Section 4 3), the function $p_{12}^r(t)$ is greater than $p_{13}^r(t)$ for all t $\in [t_0, T)$ for some t_0. Hence $\alpha_{12}^r(t)$ can not be 1 in a teminal regime. In other words, link (1,2) can not operate in Full mode in a terminal regime if link (1,3) has a linear dynamics.

EFS:

$$\begin{aligned} \dot{p}_{13}^r &= -1, \quad p_{13}^r(T) = 0 \\ \text{Hence } p_{13}^r(t) &= T - t \\ \dot{p}_{21}^r &= -1 + a_{21}(p_{21}^r - p_{13}^r), \quad p_{21}^r(T) = 0. \\ \Rightarrow \quad p_{21}^r(t) &= T - t. \\ \dot{p}_{23}^r &= -1 + a_{23}p_{23}^r, \quad p_{23}^r(T) = 0 \\ p_{23}^r(t) &= \frac{1-e^{a_{23}(t-T)}}{a_{23}} \end{aligned}$$

Since $p_{21}^r(t) = T - t > p_{23}^r(t)$, $\alpha_{21}^r(t)$ can not be unity. Thus link (2,1) can not be operate in Full mode in a terminal regime if link (1,3) is in saturation

EFL:

$$\begin{aligned} \text{Since} \quad p_{21}^r(t) &= \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{(a_{21}-a_{13})} \\ \text{is greater than} \quad p_{23}^r(t) &= \frac{1-e^{a_{23}(t-T)}}{a_{23}}, \quad \text{for all } t \text{ in some interval } [t_0, T], \end{aligned}$$

$\alpha_{21}^r(t) \not\equiv 1$ in a terminal regime. Hence **EFL** mode of operation in a terminal regime is ruled out.

Note: We proved the above result that in an interval ending with T, $p_{21}^r(t)$ is greater than $p_{23}^r(t)$, in the proof for Theorem 4 3 3

EES:

$$
\begin{aligned}
p_{13}^r(t) &= T-t \\
p_{12}^r(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)}-e^{a_{23}(t-T)}}{(a_{12}-a_{23})} \\
p_{13}^r(t) &= T-t > p_{12}^r(t) \\
\text{Therefore } \alpha_{12}^r(t) &= 1
\end{aligned}
$$

i.e. link (1,2) can not be operating in Empty mode.

FET:

It can be argued as follows that in the limiting case in which r tends to zero, the network can not operate in FET mode.

$\alpha_{12}^r(t) \equiv 1$ in an interval (denoted as I) during which the network operation is in FET mode

$$
\begin{aligned}
\text{Therefore } \dot{x}_{13}^r &= -f_{13}^r(x_{13}^r),\ \forall t \in I \\
\text{As } r \to 0,\ \dot{x}_{13} &= -C_{13}
\end{aligned}
$$

For the interval I to be a transition interval $x_{13}(t)$ has to be equal to x_{13s} $\forall t \in I$ Since this can not be satisfied if the dynamics of $x_{13}(t)$ is as given by $\dot{x}_{13} = -C_{13}$ in I (since $x_{13}(t)$ is then a monotonically decreasing function), FET mode of operation is ruled out not only in a terminal regime but also in any interval of network operation, for the case where r tends zero.

EFT:

The EFT mode of operation is also ruled out in a terminal regime for the following reasons:

Let $I = [t_0, T]$ be an interval during which the network operates in the EFT mode During I, $x_{13l} < x_{13}^r(t) < x_{13c}$ and therefore $\frac{df_{13}^r}{dx_{13}^r} < a_{13}$

$$
\text{Hence } \dot{p}_{13}^r = -1 + \frac{df_{13}^r}{dx_{13}^r} p_{13}^r \quad < -1 + a_{13} p_{13}^r \quad \forall t \in I = [t_0, T].
$$

We have the boundary condition $p_{13}^r(T) = 0$.

It can be argued that $p_{13}^r(t)$ which satisfies the above inequality in the interval I is

such that

$$p_{13}^{r}(t) > \frac{1-e^{a_{13}(t-T)}}{a_{13}}, \forall t \in [t_0, T).$$

Consider now the dynamics of $p_{21}^{r}(t)$ in I.

$$\begin{aligned} \dot{p}_{21}^{r} &= -1 + a_{21}(p_{21}^{r} - p_{13}^{r}) \\ &< -1 + a_{21}p_{21}^{r} - a_{21}\,\frac{1-e^{a_{13}(t-T)}}{a_{13}}, \quad \forall t \in [t_0, T). \end{aligned}$$

Compare $p_{21}^{r}(t)$ with $p_{21}^{*r}(t)$, where $p_{21}^{*r}(t)$ is the solution to the following differential equation:

$$\begin{aligned} p_{21}^{*r}(t) &= -1 + a_{21}[p_{21}^{*r}(t) - \frac{1-e^{a_{13}(t-T)}}{a_{13}}], \quad p_{21}^{*r}(T) = 0. \\ \text{i.e. } p_{21}^{*r}(t) &= \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{a_{21}-a_{13}} \end{aligned}$$

From the inequality,

$$p_{21}^{r}(t) - p_{21}^{*r}(t) < a_{21}(p_{21}^{r} - p_{21}^{*r}), \forall t \in [t_0, T)$$

and the boundary condition $p_{21}^{r}(T) - p_{21}^{*r}(T) = 0$, it can be argued that

$$\begin{aligned} p_{21}^{r}(t) &> p_{21}^{*r}(t), \quad \forall t \in [t_0, T) \\ \text{i e. } p_{21}^{r}(t) &> \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{a_{21}-a_{13}} \quad \forall t \in [t_0, T). \end{aligned}$$

We have already proved (in the proof for Theorem 4 3 3) in Section 4 3 that during an interval ending with T, the function $\frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)}-e^{a_{13}(t-T)}}{a_{21}-a_{13}}$ is greater than $p_{23}^{r}(t) = \frac{1-e^{a_{23}(t-T)}}{a_{23}}$.

Hence $p_{21}^{r}(t)$ is greater than $p_{23}^{r}(t)$, in an interval ending with T. Therefore $\alpha_{21}^{r}(t)$ can not be identically equal to unity during $[t_0, T]$ For the network to be operating in EFT mode during an interval, $\alpha_{21}^{r}(t)$ has to be equal to unity throughout that interval Thus EFT mode of operation is ruled out in a terminating interval.

We can rule out the modes **EPL, EPT** and **EPS** in a terminal regime from the following considerations.

EPL:

The costate variables $p_{13}^{r}(t)$, $p_{21}^{r}(t)$ and $p_{23}^{r}(t)$ are as follows

$$p_{13}^{r}(t) = -1 + a_{13}p_{13}^{r}(t), p_{13}^{r}(T) = 0.$$

$$
\begin{aligned}
\Rightarrow \; p^r_{13}(t) &= \frac{1-e^{a_{13}(t-T)}}{a_{13}} \\
p^r_{21}(t) &= \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{(a_{21}-a_{13})} \\
p^r_{23}(t) &= \frac{1-e^{a_{23}(t-T)}}{a_{23}}
\end{aligned}
$$

Link (2,1) could possibly operate in the partial mode only if $p^r_{21}(t)$ and $p^r_{23}(t)$ are identically equal over an interval. But in an interval ending with T, $p^r_{21}(t)$ as given by the above expression is greater than $p^r_{23}(t)$ (we proved this result in the proof of Theorem 4.3 3) Hence link (2,1) can not be in **Partial** mode in a terminal interval and thus **EPL** mode of operation is ruled out in a terminal regime.

EPS:

$$
\begin{aligned}
\dot{p}^r_{12} &= -1 + a_{12}(p^r_{12} - p^r_{23}) \text{ and } p^r_{12}(T) = 0 \\
p^r_{12}(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{(a_{12}-a_{23})} \\
\dot{p}_{13} &= -1 \text{ and } p^r_{13}(T) = 0. \\
p^r_{13}(t) = (T-t) &> p^r_{12}(t) \\
\text{Hence} \quad \alpha^r_{12}(t) &= 1
\end{aligned}
$$

Therefore link (1,2) can not be operating in **Empty** mode.

EPT:

Since link (1,3) is operating in the **transition** mode, $x_{13l} < x^r_{13}(t) < x_{13c}$ and consequently $\frac{df^r_{13}}{dx^r_{13}} < a_{13}$.

$$
\begin{aligned}
\dot{p}^r_{13} &< -1 + a_{13}p^r_{13}, \quad p^r_{13}(T) = 0 \\
\text{Therefore} \quad p^r_{13}(t) &> \frac{1-e^{a_{13}(t-T)}}{a_{13}} \\
\dot{p}^r_{21} &= -1 + a_{21}(p^r_{21} - p^r_{13}) \\
&< -1 + a_{21}\left(p^r_{21} - \frac{1-e^{a_{13}(t-T)}}{a_{13}}\right)
\end{aligned}
$$

As argued out earlier (in proving that **EFT** is ruled out in a terminal regime) the function $p^r_{21}(t)$ which satisfies the above inequality is such that

$$
p^r_{21}(t) > \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{e^{a_{21}(t-T)} - e^{a_{13}(t-T)}}{a_{21}-a_{13}}
$$

Since the function on the right side of the above equation is greater than $p_{23}^r(t)$ in an interval ending with T, the routing variable $\alpha_{21}^r(t)$ has to be zero Thus **EPT** mode of operation in which $\alpha_{21}^r(t)$ is non-zero is ruled out in an interval ending with T.

From the above discussion, the following conclusions can be drawn on the nature of the optimal routing strategy for networks with link (1,3) of finite channel capacity C_{13}.

- The routing variable $\alpha_{13}(t)$ takes the values 0,1 or $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}}$
- During no interval of the network operation, packets are forwarded onto both the links (1,2) and (2,1) simultaneously (**Loop-free property**)
- The network operation ends in only one of the following four possible modes
  (a) **FES :** $\alpha_{12} = 1, \alpha_{21} = 0,$ Link (1,3) in saturation
  (b) **PET :** $\alpha_{12} = 1 - \frac{C_{13}}{\lambda_1+a_{21}x_{21}}, \alpha_{21} = 0, x_{13}(t) \equiv x_{13s}.$
  (c) **EET :** $\alpha_{12} = 0, \alpha_{21} = 0, x_{13}(t) \equiv x_{13s}.$
  (d) **EEL :** $\alpha_{12} = 0, \alpha_{21} = 0,$ link (1,3) in linear mode

The examples presented below are cases wherein the network operation ends in one of the above modes

4.4.4 Numerical Examples

Consider a network with the following link parameters.

$$a_{12} = 0\,1, C_{12} = \infty.$$
$$a_{13} = 0\,5, C_{13} = 10.$$
$$a_{21} = 0\,1, C_{21} = \infty$$
$$a_{23} = 0\,4, C_{23} = \infty$$

Assume the network operation to be for a duration of 10 units. i.e T= 10 units.

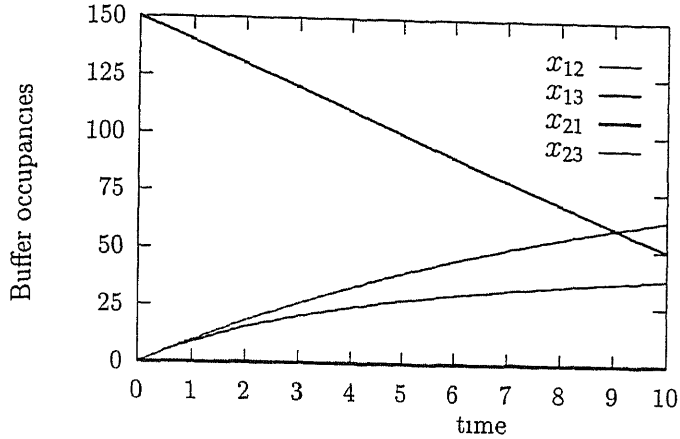

Figure 4.10 State variables as functions of time for Example 4 4.2 1

Example 4.4.2.1 :

For the following choice of the initial buffer occupancies and the input traffic $\lambda_1(t)$ and $\lambda_2(t)$ the network operation ends in **FES** mode

$$x_{13}(0) = 150, x_{12}(0) = x_{21}(0) = x_{23}(0) = 0.$$
$$\lambda_1 = \lambda_2 = 10$$

The state variables (buffer occupancies) under the optimal routing strategy are ·

$$\begin{aligned} x_{13}(t) &= 150 - 10t, \\ x_{12}(t) &= 100[1 - e^{-0\ 1t}] \\ x_{21}(t) &= 0 \\ x_{23}(t) &= 50[1 - e^{-0\ 4t}] - 33\ 33[e^{-0\ 1t} - e^{-0\ 4t}] \end{aligned}$$

The optimal routing strategy for this case is given as:

$$\alpha_{13}(t) \equiv 0, \alpha_{21}(t) \equiv 0, \forall t \in [0, 10]$$

Remark 4.4.1 It can be easily verified that the optimal routing strategy is as given above since the costate variables satisfy the following relationships·

$$\begin{aligned} p_{13}(t) &= 10 - t \\ p_{12}(t) &= 10[1 - e^{0\ 1(t-10)}] + 2\ 5[1 - e^{0\ 4(t-10)}] + 3\ 33[e^{0\ 4(t-10)} - e^{0\ 1(t-10)}] \end{aligned}$$

$$
\begin{aligned}
p_{21}(t) &= 10 - t \\
p_{23}(t) &= 2\ 5[1 - e^{0\ 4(t-10)}] \\
p_{12}(t) &< p_{13}(t),\ \forall t \in [0, 10] \ \text{ Hence } \alpha_{13}(t) \equiv 0. \\
p_{21}(t) &> p_{23}(t),\ \forall t \in [0, 10] \ \text{ Hence } \alpha_{21}(t) \equiv 0
\end{aligned}
$$

Note that this example pertains to a situation wherein the initial buffer occupancy of link (1,3) is large enough to ensure that packets can be drawn out at the maximum rate (equal to the channel capacity C_{13}) for the entire duration of network operation even when no new packets are routed onto it.

Example 4.4.2.2 :

For the following choice of initial buffer occupancies and the input traffic, the network operation ends in **PET** mode.

$$x_{13}(0) = x_{12}(0) = x_{21}(0) = x_{23}(0) = 0.$$

$$\lambda_1 = \lambda_2 = 15$$

The buffer occupancies for this case are as follows

During the interval $[0, 2\ 197]$,

$$
\begin{aligned}
x_{13}(t) &= 30(1 - e^{-0\ 5t}) \\
x_{12}(t) &= 0 \\
x_{21}(t) &= 0 \\
x_{23}(t) &= 37\ 5(1 - e^{-0\ 4t})
\end{aligned}
$$

During the interval $(2\ 197, 10]$

$$
\begin{aligned}
x_{13}(t) &= 20 \\
x_{12}(t) &= 50(1 - e^{-0\ 1(t-2\ 197)}) \\
x_{21}(t) &= 0 \\
x_{23}(t) &= 21\ 9145e^{-0\ 4(t-2\ 197)} + 50(1 - e^{-0\ 4(t-2\ 197)}) - 16\ 66(e^{-0\ 1(t-2\ 197)} - e^{-0\ 4(t-2\ 197}
\end{aligned}
$$

The optimal routing strategy for this case is given as:

$$\alpha_{13}(t) = 1, \quad \forall t \in [0, 2\ 197]$$

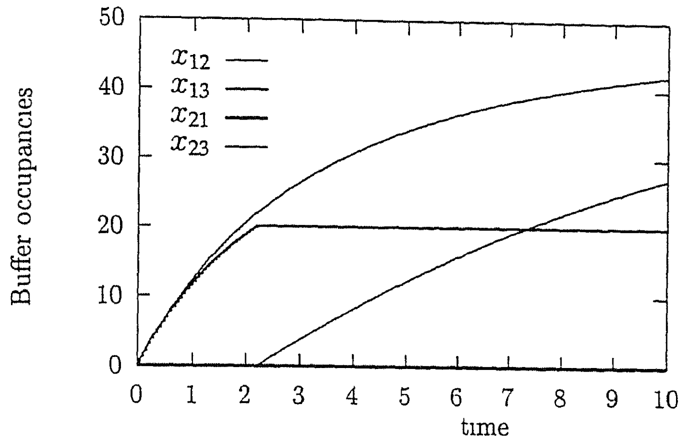

Figure 4 11 State variables as functions of time for Example 4 4 2 2

$$\begin{aligned}
\alpha_{13}(t) &= 10/15, \quad \forall t \in [2\,197, 10] \\
\alpha_{23}(t) &= 1, \quad \forall t \in [0, 10]
\end{aligned}$$

Remark 4.4.2 We shall argue in the next section that the above values of the routing variables correspond to the optimal routing strategy

Example 4.4.2.3 :

For the following choice of the initial buffer occupancies and the input traffic, the network operation ends in **EET** mode

$$x_{13}(0) = 20, x_{12}(0) = x_{21}(0) = x_{23}(0) = 0.$$
$$\lambda_1 = \lambda_2 = 10.$$

The buffer occupancies are given as:

$$\begin{aligned}
x_{13}(t) &\equiv 20 \\
x_{12}(t) &\equiv 0 \\
x_{21}(t) &\equiv 0 \\
x_{23}(t) &= 25(1 - e^{-0\,4t}) \quad \forall t \in [0, 10]
\end{aligned}$$

The optimal routing strategy is given as

$$\alpha_{12}(t) = 0, \alpha_{21}(t) = 1 \quad \forall t \in [0. 10]$$

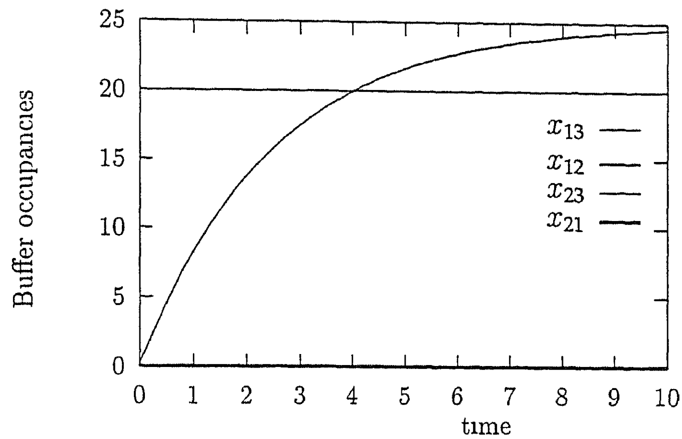

Figure 4 12 State variables as functions of time for Example 4 4.2 3 .

Remark 4.4.3 In the next section, we shall argue that the above is the optimal routing strategy.

Example 4.4.2.4 :

Consider the case wherein the initial buffer occupancies are all zero and the input traffic $\lambda_1 = \lambda_2 = 10$.

The buffer occupancies for the interval $[0, 10]$ are the following

$$\begin{aligned}
x_{13}(t) &= 20(1 - e^{-0\,5t}) \\
x_{12}(t) &= 0 \\
x_{21}(t) &= 0 \\
x_{23}(t) &= 25(1 - e^{-0\,4t})
\end{aligned}$$

The optimal routing strategy is given as:

$$\alpha_{13}(t) = \alpha_{23}(t) = 1 \quad \forall t \in [0, 10]$$

Remark 4.4.4 It can be seen that the network operates in **EEL** mode (with the above values for the routing variables) for the entire duration $[0, 10]$ since the costate variables satisfy the following relationships

$$p_{12}(t) \;=\; 10(1 - e^{0\,1(t-10)}) + 2.5(1 - e^{0\,4(t-10)}) + 3\,33(e^{0\,4(t-10)} - e^{0\,1(t-10)})$$

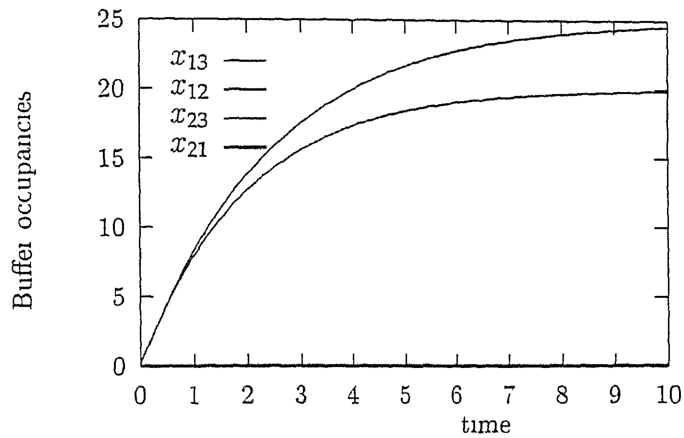

Figure 4 13. State variables as functions of time for Example 4 4.2.4

$$
\begin{aligned}
p_{13}(t) &= 2(1 - e^{0\ 5(t-10)}) \\
p_{12}(t) &> p_{13}(t) \quad \forall t \in [0, 10]. \\
p_{21}(t) &= 10(1 - e^{0\ 1(t-10)}) + 2(1 - e^{0\ 5(t-10)}) + 2\ 5(e^{0\ 4(t-10)} - e^{0\ 1(t-10)}) \\
p_{23}(t) &= 2\ 5[1 - e^{0\ 4(t-10)}] \\
p_{21}(t) &> p_{23}(t) \quad \forall t \in [0, 10].
\end{aligned}
$$

4.5 The Optimal and Suboptimal Routing Strategies Under Certain Restrictions on the Link Parameters

Our analysis in the previous section assumed no restrictions on the link parameters of the network, on the initial buffer occupancies and on the input loads $\lambda_1(t)$ and $\lambda_2(t)$ We proved some interesting properties of the optimal routing algorithm without explicitly obtaining the solutions for the routing variables. In this section, we specify the optimal routing variables for networks in which the link parameters a_{12}, a_{13}, a_{21} and a_{23} satisfy any of the three conditions of Theorem 4 3.2 stated

earlier. The implication of this assumption (regarding the link parameters) is that if all the links of the network were of infinite channel capacity (and hence the network operates is in the linear mode for the entire duration $[0,T]$) then the optimal routing strategy would be a regime 4 routing (i.e. a direct routing) over $[0,T]$. In the analysis that follows, we assume that the network operation starts in the linear mode.

The arguments used in this section for obtaining the equations for the optimal routing variables are analogous to the ones used in the case of the two-node network which we considered in Chapter 3. The optimal routing strategy is first obtained for the cases wherein the total traffic $(\lambda_1(t) + a_{21}x_{21}(t))$ arriving at node 1 has a single positive crossing over the value equal to C_{13}. The arguments are then extended to the case of any arbitrary load patterns. We derive a set of equations (in terms of the link parameters and the traffic), the solution to which specifies the optimal routing strategy. We also propose a suboptimal algorithm and provide some illustrative examples wherein the performances of the optimal and suboptimal algorithms are compared.

Assumptions:

Assumption 4.5.1 The link parameters a_{12}, a_{13}, a_{21} and a_{23} of the network satisfy any of the following sets of conditions:

(i) $a_{23} > a_{13} > a_{12}$

(ii) $a_{13} > a_{23} > a_{21}$

(iii) $a_{13} > a_{12}$ and $a_{23} > a_{21}$.

Assumption 4.5.2 The network operation starts in the linear mode i.e. $x_{13}^r(0) < x_{13l}$ (in the limiting case of r tending to zero, $x_{13}(0) < x_{13s}$).

Lemma 4.5.1 *The functions $p_{12}^r(t)$ and $p_{21}^r(t)$ are non-negative in $[0, T]$.*

Proof : It can be argued as follows that the assumption of the existence of an interval during which either $p_{12}^r(t)$ or $p_{21}^r(t)$ is negative leads to a violation of the

transversality condition

The dynamics of $p^r_{12}(t)$ and $p^r_{21}(t)$ are given as.

$$\begin{aligned} \dot{p}^r_{12}(t) &= -1 + a_{12}p^r_{12} - a_{12}\alpha^r_{21}p^r_{21} - a_{12}\alpha^r_{23}p^r_{23} \\ \dot{p}^r_{21}(t) &= -1 + a_{21}p^r_{21} - a_{21}\alpha^r_{12}p^r_{12} - a_{21}\alpha^r_{23}p^r_{23} \end{aligned}$$

Let $I_0 = [t_0, t_1]$ where $0 \leq t_0 < t_1 \leq T$ be an interval such that $p^r_{12}(t) < 0, \ \ \forall t \in I_0$. Since $p^r_{13}(t)$ is positive in the interval $[0, T)$, $p^r_{12}(t)$ is less than $p^r_{13}(t)$ in I_0

We have already proved in the previous section that there exists no interval during which $p^r_{12}(t)$ is less than $p^r_{13}(t)$ and $p^r_{21}(t)$ is less than $p^r_{23}(t)$ (see the proof of Theorem 4 4 2 in the subsection **Loop-free property**) Therefore in the interval I_0, $p^r_{21}(t)$ has to be greater than $p^r_{23}(t)$. Hence $p^r_{21}(t) > 0, \forall t \in I_0$

$$\text{Therefore,} \quad \dot{p}^r_{12} = -1 + a_{12}(p^r_{12} - p^r_{23}), \\ < -1, \quad \forall t \in I_0.$$

$p^r_{12}(t)$ is therefore a monotonically decreasing function in I_0. Hence at $t = t_1$, $p^r_{12}(t_1) < p^r_{12}(t_0) < 0$. Since the function $p^r_{12}(t)$ is continuous in the entire interval $[0, T]$, it can be concluded that I_0 is followed by an interval I_1 in which $p^r_{12}(t) < 0$ Then the above arguments (that $p^r_{12}(t)$ is monotonically decreasing and negative) can be extended to I_1, and it can be concluded that I_1, in turn, is followed by an interval in which the the function $p^r_{12}(t)$ continues to be negative and monotonically decreasing. Extension of the above arguments results in the condition $p^r_{12}(t) < 0$, $\forall t \in [t_0, T]$. Since this violates the transversality condition at $t = T$, the assumption of the existence of an interval I_0 during which $p^r_{12}(t) < 0$ is incorrect

An analogous reasoning on $p^r_{21}(t)$ can be used to prove that there exists no interval during which $p^r_{21}(t) < 0$

□

Theorem 4.5.1 *If the link parameters satisfy any of the three conditions of Assumption 4 5 1, the traffic arriving at node 2 is never routed onto the link (2,1) for the entire duration of network operation*

Proof : By Lemma 4 5 1, the functions $p^r_{12}(t)$ and $p^r_{21}(t)$ are non-negative in $[0\ T]$ We have already established (in the proof of Lemma 4.4.1) that $p^r_{13}(t)$ and $p^r_{23}(t)$ are non-negative in $[0, T]$.

$$\begin{aligned} \text{Thus,} \quad \dot{p}^r_{21} &= -1 + a_{21}p^r_{21} - a_{21}\alpha^r_{12}p^r_{12} - a_{21}\alpha^r_{13}p^r_{13} \\ &\leq -1 + a_{21}p^r_{21}, \quad \forall t \in [0, T] \end{aligned}$$

From the inequality $\dot{p}^r_{21}(t) \leq -1 + a_{21}p^r_{21}(t)\ \forall t \in [0, T]$, and the boundary condition $p^r_{21}(T) = 0$, it can be argued that

$$p^r_{21}(t) \ \geq \ \frac{(1 - e^{a_{21}(t-T)})}{a_{21}}, \forall t \in [0, T)$$

Under conditions (ii) and (iii) of Assumption 4 5 1, $a_{23} > a_{21}$

$$\begin{aligned} \text{Hence} \quad p^r_{23}(t) &< \frac{(1 - e^{a_{21}(t-T)})}{a_{21}}, \quad \forall t \in [0, T) \\ &< p^r_{21}(t), \quad \forall t \in [0, T) \\ \Rightarrow \quad \alpha^r_{21}(t) &\equiv 0, \quad \forall t \in [0, T) \end{aligned}$$

We have,

$$\begin{aligned} \dot{p}^r_{12} &\leq -1 + a_{12}p^r_{12}, \quad \text{and } p^r_{12}(T) = 0 \\ \text{Therefore} \quad p^r_{12}(t) &\geq \frac{(1 - e^{a_{12}(t-T)})}{a_{12}} \quad , \forall t \in [0, T]. \end{aligned}$$

Under condition (i) of Assumption 4.5 1, $a_{23} > a_{12}$ and $a_{23} > a_{13}$.

Therefore

$$p^r_{23}(t) < \frac{(1 - e^{a_{12}(t-T)})}{a_{12}} \leq p^r_{12}(t)$$

Furthermore

$$\dot{p}^r_{13} = -1 + \frac{df^r_{13}}{dx^r_{13}}p^r_{13} \leq -1 + a_{13}p^r_{13}, \forall t \in [0, T]$$

Hence

$$p^r_{13}(t) \geq \frac{(1 - e^{a_{13}(t-T)})}{a_{13}} > p^r_{23}(t)$$

$$\text{We know that} \quad \dot{p}^r_{21}(t) \ = \ -1 + a_{21}p^r_{21} - a_{21}\alpha^r_{12}p^r_{12} - a_{21}\alpha^r_{23}p^r_{13}$$

Since both $p^r_{12}(t)$ and $p^r_{13}(t)$ are greater than $p^r_{23}(t)$,

$$\begin{aligned}
p^r_{21}(t) &< -1 + a_{21}p^r_{21} - a_{21}p^r_{23} \quad \forall t \in [0,T) \\
\Rightarrow p^r_{21}(t) &> \frac{(1-e^{a_{21}(t-T)})}{a_{21}} + \frac{(1-e^{a_{23}(t-T)})}{a_{23}} + \frac{(e^{a_{21}(t-T)} - e^{a_{23}(t-T)})}{(a_{21}-a_{23})}, \ \forall t \in [0,T). \\
&> p^r_{23}(t), \ \forall t \in [0,T) \\
\Rightarrow \quad \alpha^r_{21}(t) &\equiv 0 \quad \forall t \in [0,T]
\end{aligned}$$

Hence under any of the three conditions (i), (ii) and (iii) of Assumption 4.5 1, $\alpha^r_{21}(t) \equiv 0$.

□

Comments :

- Under any of the three sets of conditions of Assumption 4 5.1, $p_{21}(t)$ for the linear case (i.e corresponding to the situation wherein $C_{13} = \infty$) as given by the expression

$$p_{21}(t) = \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{(e^{a_{21}(t-T)} - e^{a_{13}(t-T)})}{(a_{21}-a_{13})}$$

  is greater than $p_{23}(t)$, $\forall t \in [0,T)$, and consequently $\alpha_{21}(t) \equiv 0$. In other words, the traffic arriving at node 2 is never routed through the alternate path consisting of links (2,1) and (1,3) If we now replace link (1,3) with one of finite channel capacity (but of the same link parameter a_{13}), it is intuitively justifiable to assume that in such a network too, routing of packets arriving at node 2 along the alternate path (2,1) and (1,3) would result in a suboptimal performance. Thus the above theorem is in agreement with our intuition.

- Since $\alpha^r_{21}(t) \equiv 0, \forall t \in [0,T]$, the equation for $p^r_{12}(t)$ is

$$\begin{aligned}
\dot{p}^r_{12} &= -1 + a_{12}p^r_{12}(t) - a_{12}p^r_{23}(t) \quad p^r_{12}(T) = 0. \\
\text{Therefore} \quad p^r_{12}(t) &= \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{(a_{12}-a_{23})} \quad \text{...(D)}
\end{aligned}$$

We state below a lemma which will be used in the arguments that follow.

Lemma 4.5.2 *If $\frac{1}{a_{13}} \le \frac{1}{a_{12}} + \frac{1}{a_{23}}$, then the solution to the equation $p_{13} = -1 + a_{13}p_{13}$ with an initial condition $p_{13}(t_0)$ which satisfies the following condition*

$$p_{13}(t_0) \ge p_{12}(t_0) = \frac{1 - e^{a_{12}(t_0-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t_0-T)}}{a_{23}} + \frac{(e^{a_{12}(t_0-T)} - e^{a_{23}(t_0-T)})}{(a_{12} - a_{23})},$$

(where $t_0 < T$) is such that

$$p_{13}(t) > p_{12}(t) = \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)} - e^{a_{23}(t-T)}}{a_{12} - a_{23}}, \quad \forall t \in (t_0, T]$$

Proof : It is easily verified that $p_{12}(t)$ as given by the above expression is the solution to the following differential equation·

$$p_{12} = -1 + a_{12}p_{12} - a_{12}\frac{1 - e^{a_{23}(t-T)}}{a_{23}}$$

with the boundary condition $p_{12}(T) = 0$ We also note that

$$p_{12}(t) > \frac{1 - e^{a_{12}(t-T)}}{a_{12}} \quad \forall t \in [t_0, T]$$

If $\frac{1}{a_{13}} \le \frac{1}{a_{12}} + \frac{1}{a_{23}}$, then $\frac{1}{a_{13}} < \frac{1}{a_{12}}$ and therefore $a_{13} > a_{12}$. Similarly $a_{13} > a_{23}$
We can prove that $p_{13}(t)$ is positive in $[t_0, T]$ as follows
Compare $p_{13}(t)$ with the function $p^*_{13}(t)$ where $p^*_{13}(t)$ is the solution to the differential equation given below

$$p^*_{13}(t) = -1 + a_{13}p^*_{13}(t), \quad p^*_{13}(T) = 0$$

$$\text{i.e.} \quad p^*_{13}(t) = \frac{1 - e^{a_{13}(t-T)}}{a_{13}}$$

Since $a_{13} > a_{12}$,

$$\begin{aligned} p^*_{13}(t) &< \frac{1 - e^{a_{12}(t-T)}}{a_{12}} \quad \forall t \in [0, T) \\ &< p_{12}(t) \quad \forall t \in [0, T) \end{aligned}$$

In particular, $p^*_{13}(t_0) < p_{12}(t_0)$ Thus

$$p_{13}(t_0) \ge p_{12}(t_0) > p^*_{13}(t_0)$$

The solution to the differential equation

$$p_{13}(t) - p^*_{13} = a_{13}(p_{13}(t) - p^*_{13}(t))$$

with the initial condition $(p_{13}(t_0) - p^*_{13}(t_0)) > 0$, such that $(p_{13}(t) - p^*_{13}(t)) > 0, \forall t > t_0$. Since $p^*_{13}(t) > 0, \forall t \in [t_0, T)$, it follows that $p_{13}(t) > 0, \forall t \in [t_0, T]$. This proves that the function $p_{13}(t)$ is positive in $[t_0, T]$.

Thus under the stated inequality of the lemma, we have the following relationships,

$$
\begin{aligned}
p_{13} &= -1 + a_{13} p_{13} \\
&> -1 + a_{12} p_{13} \quad \text{since } p_{13}(t) \text{ is positive.} \\
p_{12} &= -1 + a_{12} p_{12} - a_{12} \frac{1 - e^{a_{23}(t-T)}}{a_{23}} \\
&< -1 + a_{12} p_{12} \\
\text{Therefore,} \quad p_{13} - p_{12} &> a_{12}(p_{13} - p_{12}) \forall t \in [t_0, T]
\end{aligned}
$$

From the initial condition $p_{13}(t_0) - p_{12}(t_0) \geq 0$, and from the above inequality, it follows that $p_{13}(t) > p_{12}(t), \forall t \in [t_0, T]$
Hence the lemma.

□

Theorem 4.5.2 *The routing variable $\alpha_{13}(t)$ (in the limiting case of r tending to zero) is either 1 or $\frac{C_{13}}{\lambda_1(t) + a_{21} x_{21}(t)}$, $\forall t \in [0, T]$*

Proof : We prove that for the control problem corresponding to the limiting case of r tending to zero, there exists no interval $I = [t_0, t_1]$ during which $p^r_{13}(t) > p^r_{12}(t)$.
Let $p^r_{13}(t) > p^r_{12}(t), \quad \forall t \in I = [t_0, t_1]$.
Then $\alpha^r_{13}(t) \equiv 0, \quad \forall t \in I$, and $x^r_{13} = -f^r_{13}(x^r_{13})$ Therefore $x^r_{13}(t)$ is a monotonically decreasing function in I.

It can be inferred that $x^r_{13}(t) < x_{13c}, \forall t \in I$ from the following arguments.
If $x^r_{13}(t) \geq x_{13c}$ for any t_i, then $x^r_{13}(t) > x_{13c}, \forall t \in [t_0, t_i]$ (since $x^r_{13}(t)$ is monotonically decreasing in I) Consider the following set of differential equations in the state variable $x^r_{13}(t)$ and the costate variable $p^r_{13}(t)$ in the interval $[0, t_i]$.

$$
\begin{aligned}
x^r_{13} &= -f^r_{13}(x^r_{13}) + \alpha^r_{13}(t)(\lambda_1(t) + a_{21} x^r_{21}) \\
p^r_{13} &= -1 + \frac{d f^r_{13}(x^r_{13})}{d x^r_{13}} p^r_{13}
\end{aligned}
$$

where $\alpha_{13}^r(t) = 0$ if $p_{13}^r(t) > p_{12}^r(t)$ (which is given by the Equation (D)). It is easy to verify that if the boundary conditions are such that $x_{13}^r(t_i) > x_{13c}$ and $p_{13}^r(t_i) > p_{12}^r(t_i)$, then the solution to $x_{13}^r(t)$ and $p_{13}^r(t)$ are given as

$$\begin{aligned} x_{13}^r(t) &= x_{13}^r(t_i) + C_{13}(t_i - t) \text{ where } x_{13}^r(t_i) > x_{13c} \\ p_{13}^r(t) &= p_{13}^r(t_i) + (t_i - t) \end{aligned}$$

Thus at $t = 0$, we get $x_{13}^r(0) > x_{13c} + C_{13}(t_i)$ which violates the Assumption 4 5 2 that the network operation starts in the linear mode Therefore $x_{13}^r(t)$ has to be less than x_{13c} during $[t_0, t_i]$.

It can also be concluded that $x_{13}^r(t)$ is greater than x_{13l}, $\forall t \in I$ from the following arguments.

If $x_{13}^r(t) \leq x_{13l}$ for some $t_i \in I$, then $x_{13}^r(t) < x_{13l}$, $\forall t > t_i$ and $\dot{p}_{13}^r(t) = -1 + a_{13}p_{13}^r$, $\forall t \in [t_i, t_1]$ Note that the link parameters a_{13}, a_{12} and a_{23} satisfy the condition of Lemma 4 5 2 Therefore the function $p_{13}^r(t)$ is greater than $p_{12}^r(t)$ in the entire interval $[t_i, T]$ Since this results in the violation of the transversality condition $p_{13}^r(T) = 0$, it can be concluded that during the interval I, $x_{13}^r(t) > x_{13l}$.

Thus, $x_{13l} < x_{13}^r(t) < x_{13c}$.

Therefore

$$\begin{aligned} \dot{x}_{13}^r &< -a_{13}x_{13l}, \quad \forall t \in I. \\ \text{which implies that} \quad x_{13}^r(t_1) - x_{13}^r(t_0) &< -a_{13}x_{13l}(t_1 - t_0) \\ \text{i.e} \quad x_{13}^r(t_0) - x_{13}^r(t_1) &> a_{13}x_{13l}(t_1 - t_0) \end{aligned}$$

Since $(x_{13}^r(t_0) - x_{13}^r(t_1)) < x_{13c} - x_{13l}$, (both $x_{13}^r(t_0)$ and $x_{13}(t_1)$ lie in the interval $[x_{13l}, x_{13c}]$) we get the inequality:

$$0 < (t_1 - t_0) < \frac{(x_{13c} - x_{13l})}{a_{13}x_{13l}}$$

In the limiting case as r tends to zero, both x_{13c} and x_{13l} tend to x_{13s}. Hence the length of the interval I which is equal to $(t_1 - t_0)$ also tends to zero In other words, corresponding to the control problem corresponding to the limiting case of r tending to zero, there doesn't exist an interval during which $p_{13}(t) > p_{12}(t)$

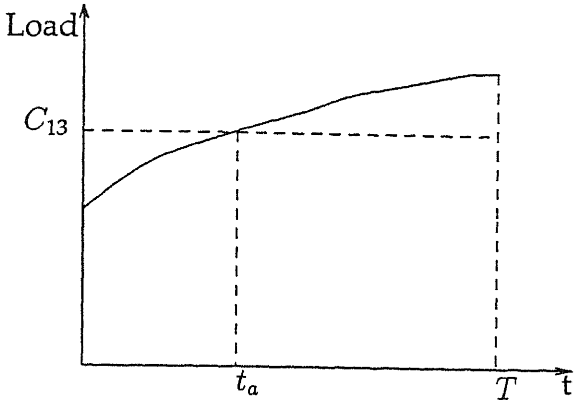

Figure 4.14: Load pattern which ends with a value greater than C_{13}

We have already established that if $p_{13}(t)$ is less than $p_{12}(t)$, then $\alpha_{13}(t)$ equals unity and if $p_{13}(t)$ is identically equal to $p_{12}(t)$ over an interval then $\alpha_{13}(t)$ equals $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$ during that interval. Thus the only possible values $\alpha_{13}(t)$ can take are 1 and $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$ Hence the theorem

□

It can be easily argued that an *interval of partial routing* during which $p_{12}(t)$ is identically equal to $p_{13}(t)$ (and $\alpha_{13}(t)$ equals $\frac{C_{13}}{(\lambda_1(t)+a_{21}x_{21}(t))}$) can not be preceeded by a saturation regime ($p_{13} = -1$) nor be followed by a linear regime ($p_{13} = -1+a_{13}p_{13}$) if the Assumptions 4 5.1 and 4 5 2 are satisfied since either of the above results in an interval during which $p_{13}(t) > p_{12}(t)$.

We now examine the nature of the optimal routing strategy for the case in which the total traffic arriving at node 1 ($\lambda_1(t) + a_{21}x_{21}(t)$) has a single positive crossing over a value equal to C_{13} as shown in Figure 4 14 and Figure 4.15

Note: The arguments that follow are applicable to the case of r tending to zero only

As proved in the previous section, the network operation can end only in one of the four modes . **FES, PET, EET** and **EEL**. For the network operation to end in **FES**, it is necessary (and sufficient) that $x_{13}(0) > x_{13s} + C_{13}T$ Since this is not the case as per the Assumption 4 5 2, **FES** as a terminal mode is ruled out

A necessary condition for the network to end in **PET** or **EET** is that $\lambda_1(t)+a_{21}x_{21}(t)$

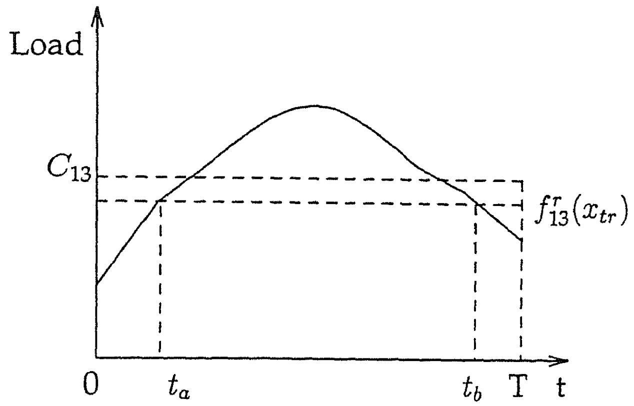

Figure 4.15 Load $\lambda_1(t) + a_{21}x_{21}(t)$ wıth a sıngle positıve crossıng over the value C_{13}

ıs greater than or equal to C_{13}, $\forall t \in [t_0, T]$ for some $0 \le t_0 < T$ For loads shown ın Fıgure 4 15 such ıs not the case and therefore the network operatıon ends ın **EEL** mode.

Consıder loads whıch end wıth a value greater than C_{13} as shown ın Figure 4 14 By the Theorem 4 5.2 the routıng variable $\alpha_{13}(t)$ ıs eıther unıty or $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}}$ If $x_{13}(t) = x_{13s}$ at some ınstant $t_0 > t_a$ (where t_a ıs as shown ın the Fıgure 4.14) then

$$\alpha_{13}(t) = \frac{C_{13}}{\lambda_1(t) + a_{21}x_{21}(t)} \quad \text{for } [t_0, T].$$

Note Thıs follows from the fact that the network operatıon can not end ın FES mode

Therefore the optımal routıng strategy for loads $(\lambda_1(t) + a_{21}x_{21}(t))$ which end wıth a value greater than C_{13}, is

$$\begin{aligned}
\alpha_{13}(t) &= 1 \quad \text{For } t \in [0, t_s] \\
&= \frac{C_{13}}{\lambda_1(t) + a_{21}x_{21}(t)} \quad \text{for } t \in [t_s, T] \\
\alpha_{12}(t) &= 1 - \alpha_{13}(t) \\
\text{and} \quad \alpha_{21}(t) &\equiv 0, \\
\alpha_{23}(t) &\equiv 1
\end{aligned}$$

where t_s is the first ınstant at whıch $x_{13}(t)$ reaches the value x_{13s}.

4.5.1 Optimal routing strategy: Implementation

We have already established in Theorem 4 5.2 that either the entire traffic at node 1 is routed onto the link (1,3) or there is a *partial routing* of $\frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$ onto this link. By Theorem 4.5.1 the routing variable $\alpha_{21}(t)$ is zero for the entire interval of network operation. Therefore the optimal routing strategy for this network under the stated Assumptions 4 5 1 and 4.5 2 can be implemented by the following algorithm·

Step I:

Check if $\alpha_{13}(t) = 1, \forall t \in [0,T]$ is the optimal routing strategy by the following steps

(i) Integrate the equations $x_{13} = -f_{13}(x_{13}) + \lambda_1(t) + a_{21}x_{21}(t)$, where $x_{21}(t) = x_{21}(0)e^{-a_{21}t}(\alpha_{21}(t) \equiv 0$ and therefore $x_{21} = -a_{21}x_{21})$, with the initial condition $x_{13}(0)$. Obtain $x_{13}(t)$ and $\frac{df_{13}}{dx_{13}}, \forall t \in [0,T]$.

(ii) Check if the following inequality holds true

$$p_{13}(t) < p_{12}(t) = \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{e^{a_{12}(t-T)}-e^{a_{23}(t-T)}}{(a_{12}-a_{23})}$$

If the inequality is satisfied for all $t \in [0,T]$ then $\alpha_{13}(t) \equiv 1$ is the optimal routing strategy.

Step II.

If the inequality in step I (ii) is violated for some $t \in [0,T]$, there are *intervals of partial routing* during which

$$\alpha_{13}(t) = \frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$$

We shall investigate a procedure to specify the *intervals of partial routing* if they exist Consider the case where in the total load arriving at node 1 has a single positive crossing over the value equal to C_{13}. Since an *interval of partial routing* can't be preceded by a saturation regime nor be followed by a linear regime it has to be $[t_s, t_1]$ where t_s is the first instant when $x_{13}(t) = x_{13s}$, and t_1 is such that $t_a < t_s < t_1 < t_b$ Here the instants t_a and t_b as shown in the Figure 4.15 correspond to the instants at which the load pattern crosses the value C_{13}.

The optimal routing strategy is therefore given by

$$\begin{aligned}
\alpha_{13}(t) &= 1 \quad \forall t \in [0, t_s]. \\
&= \frac{C_{13}}{\lambda_1(t) + a_{21}x_{21}(t)} \quad \forall t \in [t_s, t_1]. \\
&= 1 \quad \forall t \in (t_1, T]. \\
\alpha_{12}(t) &= 1 - \alpha_{13}(t) \\
\alpha_{21}(t) &\equiv 0 \\
\alpha_{23}(t) &\equiv 1.
\end{aligned}$$

Solution for t_1:

During $[t_s, t_1]$, $p_{12}(t) \equiv p_{13}(t)$.

Hence $$p_{13}(t_1) = p_{12}(t_1) = \frac{1 - e^{a_{12}(t_1 - T)}}{a_{12}} + \frac{1 - e^{a_{23}(t_1 - T)}}{a_{23}} + \frac{e^{a_{12}(t_1 - T)} - e^{a_{23}(t_1 - T)}}{(a_{12} - a_{23})}$$

At t_1, $x_{13}(t_1) = x_{13s}$

Let t_1^* be the instant ($t_1^* > t_1$) when $x_{13}(t)$ reaches the value x_{13s}. During $[t_1, t_1^*]$, $x_{13}(t) > x_{13s}$.

$$\int_{t_1}^{t_1^*} (\lambda_1(t) + a_{21}x_{21}(t))dt = C_{13}(t_1^* - t_1) \tag{4 32}$$

The dynamics of $p_{13}(t)$ over $[t_1, t_1^*]$ is given by

$$p_{13} = -1$$

Hence for $t \in [t_1, t_1^*]$,

$$\begin{aligned}
p_{13}(t) &= p_{13}(t_1) - (t - t_1) \\
&= \frac{1 - e^{a_{12}(t_1 - T)}}{a_{12}} + \frac{1 - e^{a_{23}(t_1 - T)}}{a_{23}} + \frac{e^{a_{12}(t_1 - T)} - e^{a_{23}(t_1 - T)}}{a_{12} - a_{23}} - (t - t_1)
\end{aligned}$$

Over the interval $[t_1^*, T]$, link (1,3) is in the linear mode Therefore

$$p_{13}(t_1^*) = \frac{1 - e^{a_{13}(t_1^* - T)}}{a_{13}}$$

Hence

$$\frac{1 - e^{a_{12}(t_1 - T)}}{a_{12}} + \frac{1 - e^{a_{23}(t_1 - T)}}{a_{23}} + \frac{e^{a_{12}(t_1 - T)} - e^{a_{23}(t_1 - T)}}{(a_{12} - a_{23})} - (t_1^* - t_1) = \frac{1 - e^{a_{13}(t_1^* - T)}}{a_{13}} \tag{4 33}$$

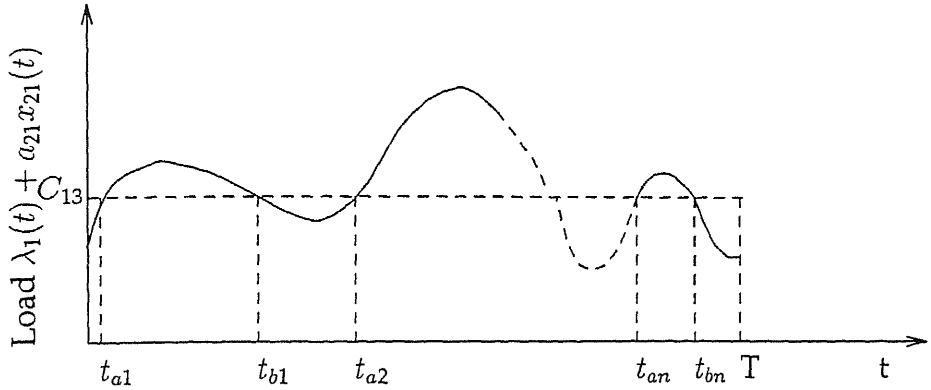

Figure 4.16: Load $\lambda_1(t) + a_{21}x_{21}(t)$ which has n positive crossings over the value C_{13}

Solving the Equations (4 32) and (4 33) we get the instants t_1 and t_1^*

Comments: The arguments used for obtaining Equations (4.32) and (4 33) are similar to those used for obtaining Equations (3 16) and (3 17) in the Chapter 3 for the two node case.

4.5.2 Extension to arbitrary load patterns

Consider a load pattern $\lambda_1(t)$ which is such that $\lambda_1(t) + a_{21}x_{21}(t)$ has n positive crossings above the value C_{13} as shown in the Figure 4.16

Notes:

(1) In the arguments that follow, we consider the case where r tends to zero

(2) The analysis below is along similar lines as in the two-node case of Chapter 3.

From our earlier discussion on the nature of the Optimal Routing Strategy, it can be concluded that *intervals of partial routing* if they exist occur in the time intervals during which $\lambda_1(t) + a_{21}x_{21}$ is above the value C_{13} Since they are never preceded by saturation regimes nor followed by linear regimes, they must be of the type $[t_{s1}, t_1]$, $[t_{s2}, t_2]$, $[t_{sn}, t_n]$ where $t_{ai} \le t_{si} \le t_i < t_{bi}$ and t_{s1}, t_{s2} are the instants at which $x_1(t) = x_{1s}$.

The optimal routing strategy is specified as.

$$\begin{aligned}
\alpha_{13}(t) &= \frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)} \text{ during the interval } [t_{si}, t_i], \text{ i=1, 2, 3, ., n.}\\
&= 1 \quad \text{everywhere else}\\
\alpha_{12}(t) &= 1-\alpha_{13}(t),\\
\alpha_{21}(t) &\equiv 0\\
\alpha_{23}(t) &\equiv 1
\end{aligned}$$

Let (t_i, t_i^*) be the saturation regime which follows the i^{th} interval of partial routing (since an interval of partial routing can not be followed by a linear regime, it has to be followed by a saturation regime during which $x_{13}(t)$ is greater than x_{13s})
At $t = t_i^*$, $x_{13}(t_i^*) = x_{13}(t_i) = x_{13s}$ for i=1, 2, .., n.
Then

$$\int_{t_i}^{t_i^*}(\lambda_1(t)+a_{21}x_{21}(0)e^{-a_{21}t})dt = C_{13}(t_i^*-t_i), \quad \text{for i=1, 2, ..n.} \tag{4.34}$$

During the interval $[t_i^*, t_{s_{i+1}}]$, $\alpha_{13}(t) = 1$ and the network operation is in the linear mode. Therefore the dynamics of link (1,3) is given by

$$x_{13} = -a_{13}x_{13} + (\lambda_1(t)+a_{21}x_{21}(t)), \text{ and } x_{13}(t_i^*) = x_{13s}.$$

Therefore

$$x_{13}(t_{s_{i+1}}) = x_{13s}e^{-a_{13}(t_{s_{i+1}}-t_i^*)} + e^{-a_{13}t_{s_{i+1}}}\int_{t_i^*}^{t_{s_{i+1}}} e^{a_{13}t}(\lambda_1(t)+a_{21}x_{21}(t))dt$$
$$\text{for i=1, 2, 3, ..., (n-1).}$$

Since $x_{13}(t_{s_{i+1}}) = x_{13s}$, we get the following (n-1) equations in t_{s_i}'s and t_i^*'s

For i= 1, 2, . .., (n-1)

$$x_{13s}(1-e^{-a_{13}(t_{s_{i+1}}-t_i^*)}) = e^{-a_{13}t_{s_{i+1}}}\int_{t_i^*}^{t_{s_{i+1}}} e^{a_{13}t}[\lambda_1(t)+a_{21}x_{21}(0)e^{-a_{21}t}]dt \tag{4.35}$$

During the interval $[t_i, t_i^*]$, $x_{13}(t) > x_{13s}$ and therefore

$$\begin{aligned}
p_{13} &= -1,\\
\text{with } p_{13}(t_i) = p_{12}(t_i) &= \frac{1-e^{a_{12}(t_i-T)}}{a_{12}} + \frac{1-e^{a_{23}(t_i-T)}}{a_{23}} + \frac{e^{a_{12}(t_i-T)}-e^{a_{23}(t_i-T)}}{(a_{12}-a_{23})}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
p_{13}(t_i) &= p_{13}(t_i^*) + (t_i^* - t_i) \\
\text{i.e. } (t_i^* - t_i) + p_{13}(t_i^*) &= \frac{1-e^{a_{12}(t_i - T)}}{a_{12}} + \frac{1-e^{a_{23}(t_i - T)}}{a_{23}} + \frac{e^{a_{12}(t_i - T)} - e^{a_{23}(t_i - T)}}{(a_{12} - a_{23})}
\end{aligned} \tag{4.36}
$$

Over $[t_i^*, t_{s_i+1}]$, link (1,3) operates in the linear mode and therefore the dynamics of $p_{13}(t)$ is given as

$$
\begin{aligned}
\dot{p}_{13} &= -1 + a_{13}p_{13} \\
\text{Since } p_{13}(t_{s_i+1}) &= p_{12}(t_{s_i+1}), \quad \text{we get} \\
p_{13}(t_i^*) &= p_{12}(t_{s_i+1})e^{a_{13}(t_i^* - t_{s_i+1})} + \frac{1-e^{-a_{13}(t_{s_i+1} - t_i^*)}}{a_{13}}
\end{aligned}
$$

Substituting the above in the Equation (4 36), we get the following n equations For i= 1, 2, .. , n

$$
\left.\begin{array}{l}
\frac{1-e^{a_{12}(t_i - T)}}{a_{12}} + \frac{1-e^{a_{23}(t_i - T)}}{a_{23}} \\
+\frac{e^{a_{12}(t_i - T)} - e^{a_{23}(t_i - T)}}{(a_{12} - a_{23})}
\end{array}\right\} = \left\{\begin{array}{l}
(t_i^* - t_i) + \frac{1-e^{-a_{13}(t_{s_i+1} - t_i^*)}}{a_{13}} + \\
\left(\frac{1-e^{a_{12}(t_{s_i+1} - T)}}{a_{12}} + \frac{1-e^{a_{23}(t_{s_i+1} - T)}}{a_{23}}\right. \\
\left. + \frac{e^{a_{12}(t_{s_i+1} - T)} - e^{a_{23}(t_{s_i+1} - T)}}{(a_{12} - a_{23})}\right) e^{a_{13}(t_i^* - t_{s_i+1})}
\end{array}\right. \tag{4 37}
$$

To solve for $t_1, t_1^*, t_{s_2}, t_2, t_2^*, \ldots t_{s_n}, t_n, t_n^*$, we have to solve the (3n-1) Equations (4.34), (4 35), and (4.37) given above.

Analytical solutions to the above set of equations are not easy to obtain. Furthermore, the specification of the optimal routing strategy requires the knowledge of the input load $\lambda_1(t)$ for the entire duration $[0, T]$ Since this implies that an on-line implementation of the routing strategy cannot be obtained, we propose the following suboptimal strategy

4.5.3 Suboptimal Algorithm

Since the maximum rate at which link (1,3) can be operated is its channel capacity C_{13} which is achieved when the buffer occupancy of this link reaches a value equal

to x_{13s}, it would be advantageous to route the overflow traffic via link (1,2) and (2,3) to the destination node 3, once the buffer occupancy of link (1,3) reaches the value x_{13s} Based on this intuition (which was also the basis for the suboptimal algorithm for the two-node case of Chapter 3) we consider the following suboptimal strategy

$$\begin{aligned}
\alpha_{13}(t) &= 1 \quad \text{for} \quad t \in [0, t_{s1}] \\
&= \frac{C_{13}}{\lambda_1(t) + a_{21}x_{21}(t)} \text{for} \quad t \in [t_{s_1}, t_{b_1}] \\
&= 1 \quad \text{everywhere else} \\
\text{and} \quad \alpha_{12}(t) &= 1 - \alpha_{13}(t) \\
\alpha_{21}(t) &\equiv 0 \quad \text{over } [0, T] \\
\alpha_{23}(t) &\equiv 1 \quad \text{over } [0, T].
\end{aligned}$$

For some typical load patterns, the performances of the optimal strategy and the suboptimal strategy are compared in the examples given below.

4.5.4 Numerical Examples

Consider a network with parameters $a_{12} = 0\ 1$, $a_{13} = 0\ 5$, $a_{21} = 0\ 1$, $a_{23} = 0\ 5$, and $C_{13} = 5\ 0$, with initial buffer occupancies $x_{13}(0) = x_{12}(0) = x_{21}(0) = x_{23}(0) = 0$ Let the network operation be for a time span $[0, 10]$ The buffer occupancies, routing variables and the performances for the optimal and the suboptimal strategies (which is proposed above) are given below for each of the example.

Note :

In each of the five examples given below, we assume that the input traffic at node 2 (i e. $\lambda_2(t)$) is a constant equal to 10 for the entire duration of network operation.

Example 4.5.4.1 :

The input traffic at node 1 .i e $\lambda_1(t)$ is 10 for the entire duration $[0, 10]$ as shown in Figure 4 17. For this case the optimal and the suboptimal strategies are identical.

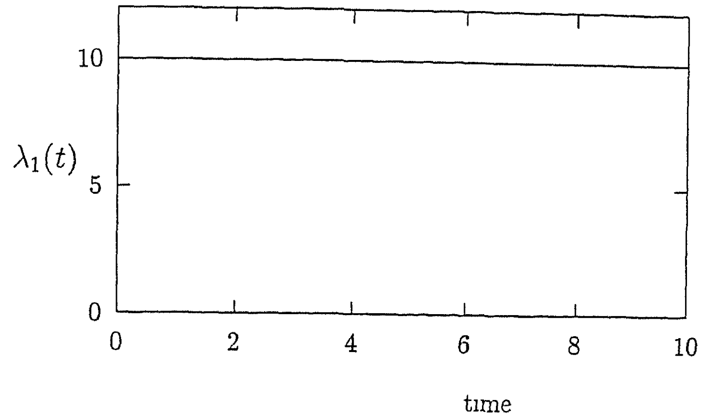

Figure 4.17 Load $\lambda_1(t)$ for Example 4 5 4.1

Table 4 1 Optimal and suboptimal routing strategies for Example 4 5 4.1.

Optimal Strategy	Suboptimal Strategy
$\alpha_{13}(t) = 1$ for $t \in [0, 1\ 3863]$ $= 0\ 5, t \in (1\ 3863, 10]$	$\alpha_{13}(t) = 1$ for $t \in [0, 1\ 3863]$ $= 0\ 5, t \in (1\ 3863, 10]$
$\alpha_{23}(t) = 1$ for $t \in [0, 10]$	$\alpha_{23}(t) = 1$ for $t \in [0, 10]$
$J_{opt} = 375\ 00126$	$J_{subopt} = 375\ 00126$

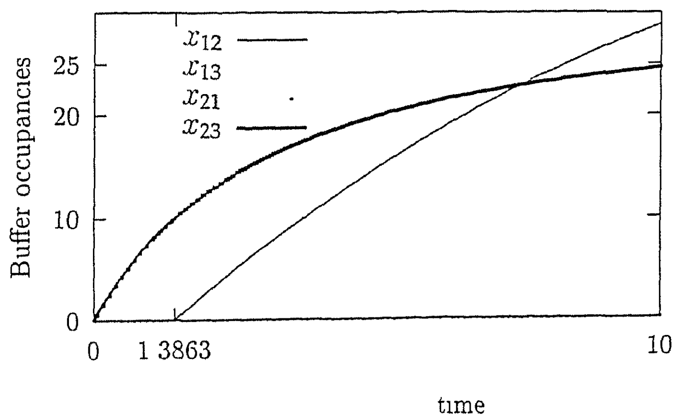

Figure 4.18 Buffer occupancies under the optimal and suboptimal strategies for Example 4 5 4 1.

Example 4.5.4.2 :

The load pattern $\lambda_1(t)$ is as shown below:

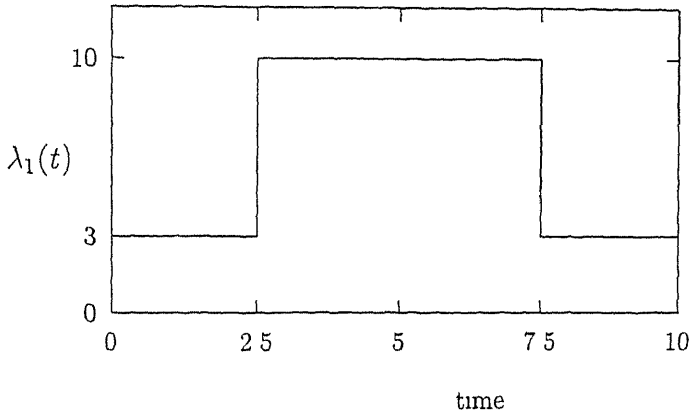

Figure 4.19: Load $\lambda_1(t)$ for Example 4.5 4 2

Table 4 2. Optimal and suboptimal routing strategies for Example 4.5.4 2.

Optimal Strategy	Suboptimal Strategy
$\alpha_{13}(t) = 1$ for $t \in [0, 3\ 4046]$ $= 0\ 5, t \in (3.4046, 6.852491]$ $=1$ for $t \in (6\ 852491, 10]$	$\alpha_{13}(t) = 1$ for $t \in [0, 3\ 4046]$ $=0\ 5, t \in (3\ 4046, 7\ 5]$ $=1, t \in (7\ 5, 10]$
$\alpha_{23}(t) = 1$ for $t \in [0, 10]$	$\alpha_{23}(t) = 1$ for $t \in [0, 10]$
$J_{opt} = 330\ 24194$	$J_{subopt} = 334\ 77986$

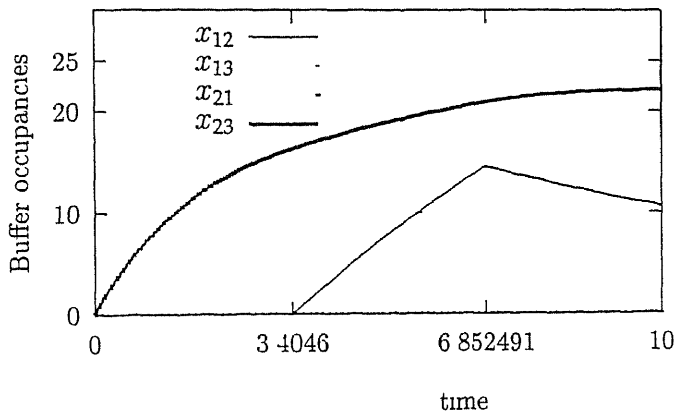

Figure 4.20 Buffer occupancies under the optimal strategy for Example 4 5.4.2.

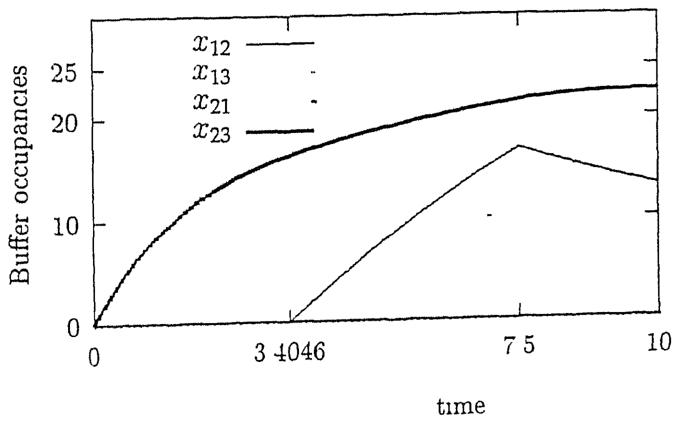

Figure 4.21 Buffer occupancies under the suboptimal strategy for Example 4.5 4 2.

Example 4.5.4.3 :

The load pattern $\lambda_1(t)$ is as shown below:

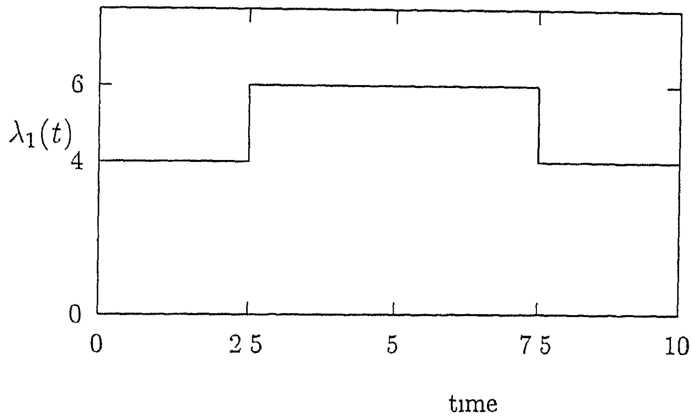

Figure 4 22. Load $\lambda_1(t)$ for Example 4 5 4.3.

Table 4 3· Optimal and suboptimal routing strategies for Example 4 5 4 3.

Optimal Strategy	Suboptimal Strategy
$\alpha_{13}(t) = 1$ for $t \in [0, 4\ 79228]$ $= 5/6, t \in (4\ 79228, 6\ 1617]$ $= 1, t \in (6\ 1617, 10]$	$\alpha_{13}(t) = 1$ for $t \in [0, 4\ 79228]$ $= 5/6, t \in (4\ 79228, 7\ 5]$ $= 1, t \in (7\ 5, 10]$
$\alpha_{23}(t) = 1$, for $t \in [0, 10]$	$\alpha_{23}(t) = 1$, for $t \in [0, 10]$
$J_{opt} = 249\ 01467$	$J_{subopt} = 250\ 6276$

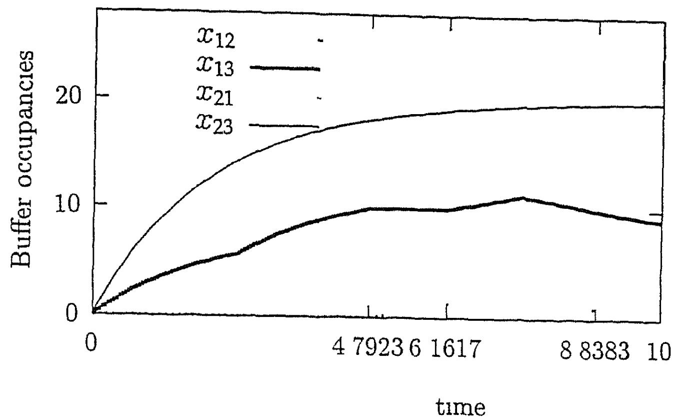

Figure 4.23: Buffer occupancies under the optimal strategy for Example 4.5.4.3

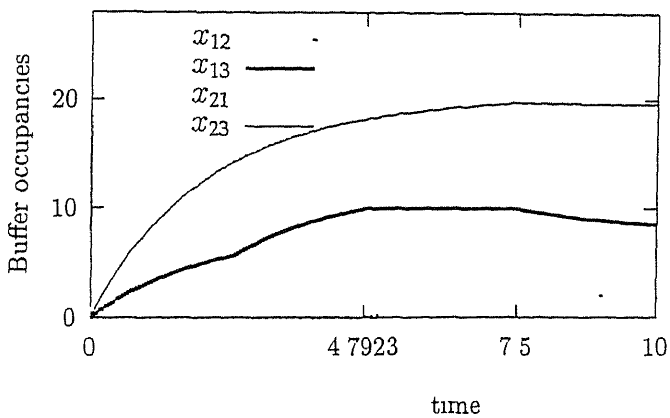

Figure 4 24 Buffer occupancies under the suboptimal strategy for Example 4 5.4 3

Example 4.5.4.4 :

The load pattern $\lambda_1(t)$ is as shown below·

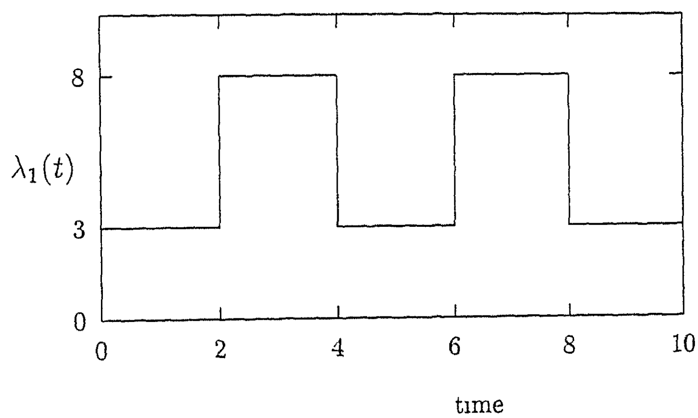

Figure 4 25 Load $\lambda_1(t)$ for Example 4 5 4 4

Table 4.4 Optimal and suboptimal routing strategies for Example 4 5 4.4.

Optimal Strategy	Suboptimal Strategy
$\alpha_{13}(t) = 1$ for $t \in [0, 6\ 5060493]$ $= 5/8$ over $(6\ 5060, 6\ 710]$ $= 1$ over $(6\ 710, 10]$	$\alpha_{13}(t) = 1$ for $t \in [0, 3\ 4205455]$ $= 5/8$, $t \in (3\ 4205455. 4]$ $=1$,$t \in (4, 6\ 70330]$ $=5/8$,$t \in (6\ 70330. 8]$ $=1$, $t \in (8, 10]$
$\alpha_{23}(t) = 1$, for $t \in [0, 10]$	$\alpha_{23}(t) = 1$, for $t \in [0. 10]$
$J_{opt} = 249\ 43406$	$J_{subopt} = 255\ 5198$

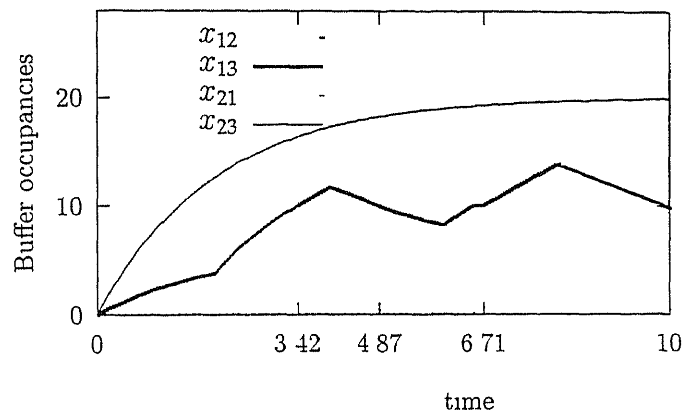

Figure 4.26 Buffer occupancies under the optimal strategy for Example 4 5 4 4.

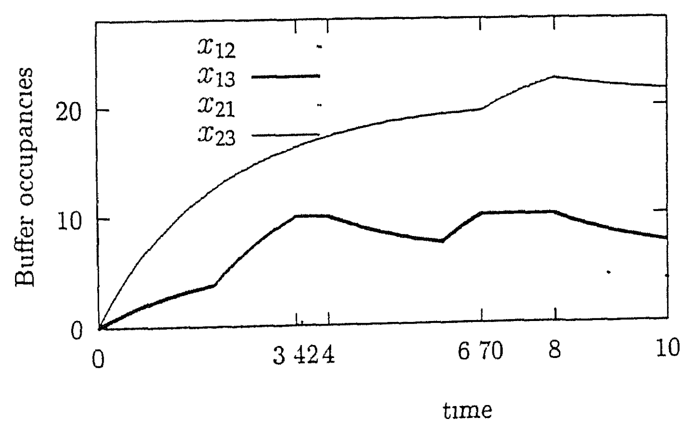

Figure 4 27 Buffer occupancies under the suboptimal strategy for Example 4 5 4 4

Example 4.5.4.5 :

The load pattern $\lambda_1(t)$ is as shown below·

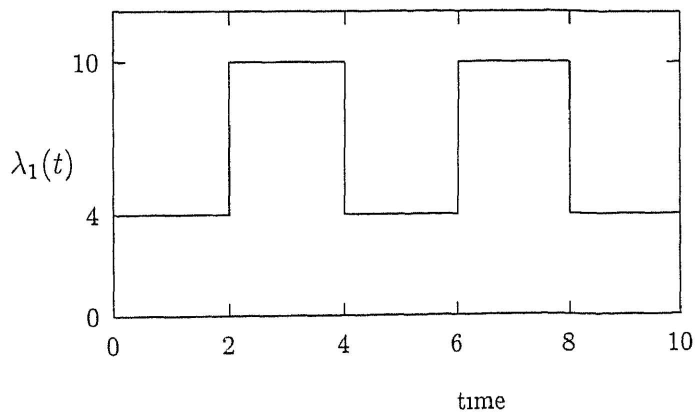

Figure 4 28: Load $\lambda_1(t)$ for Example 4.5.4 5

Table 4 5 Optimal and suboptimal routing strategies for Example 4 5.4 5.

Optimal Strategy	Suboptimal Strategy
$\alpha_{13}(t) = 1$ for $t \in [0, 2\ 8033205]$ $= 0\ 5, t \in (2.8033205, 3\ 655]$ $= 1$ for $t \in (3\ 655, 6\ 0507372]$ $=0\ 5$ for $t \in (6\ 0507372, 7\ 712575]$ $= 1$ for $t \in (7\ 712575\ \ 10]$	$\alpha_{13}(t) = 1$ for $t \in [0, 2\ 8033205$ $=0\ 5, t \in (2\ 8033205, 4]$ $=1, t \in (4\ \ 6\ 2380962]$ $=0\ 5, t \in (6\ 2380962, 8]$ $= 1$ for $t \in [8, 10]$
$\alpha_{23}(t) = 1$, for $t \in [0, 10]$	$\alpha_{23}(t) = 1$, for $t \in [0, 10]$
$J_{opt} = 291\ 52963$	$J_{subopt} = 295\ 02552$

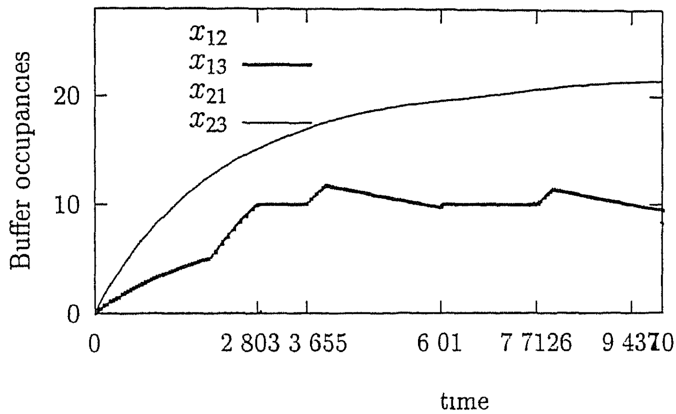

Fıgure 4 29 Buffer occupancıes under the optımal strategy for Example 4 5.4 5

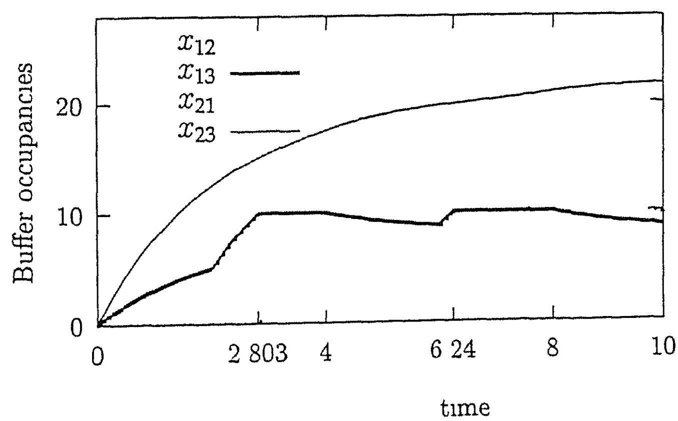

Fıgure 4 30 Buffer occupancıes under the suboptımal strategy for Example 4.5.4.5

Finally, we make the following observations regarding the nature of the optimal routing strategy for the case where the network operation starts with an initial buffer occupancy $x_{13}(0) \geq x_{13s}$ If the initial buffer occupancy $x_{13}(0)$ is s t. $x_{13}(0) > x_{13s} + C_{13}T$, then $\alpha_{13}(t) \equiv 0$, and $\alpha_{21}(t) \equiv 0$ is the optimal routing strategy for the entire duration $[0, T]$. For initial buffer occupancies $x_{13}(0)$ s t $x_{13s} < x_{13}(0) < x_{13s} + C_{13}T$, it can be easily argued that the network operation can not end in saturation mode and hence ends in either the linear mode or the transition mode It can also be argued that if $x_{13}(t)$ at any instant t_0 is equal to x_{13s}, and if $\lambda_1(t) + a_{21}x_{21}(t) > C_{13}$ for all $t > t_0$, then $\alpha_{13}(t) = \frac{C_{13}}{\lambda_1(t)+a_{21}x_{21}(t)}$ for all $t \in [t_0, T]$ It is easy to see the analogy of all the above results with those of the two-node case of Chapter 3.

4.6 Exponential Model : Some Observations

Another interesting class of models which represents a flowout function that increases with increasing buffer occupancy and saturates at the channel capacity value of the associated link is the exponential model given below·

$$
\begin{aligned}
x_{13} &= -C_{13}(1 - e^{-\frac{a_{13}x_{13}}{C_{13}}}) + \alpha_{13}(\lambda_1(t) + C_{21}(1 - e^{-a_{21}x_{21}/C_{21}})) && (4.38)\\
x_{12} &= -C_{12}(1 - e^{-a_{12}x_{12}/C_{12}}) + \alpha_{12}(t)(\lambda_1(t) + C_{21}(1 - e^{-a_{21}x_{21}/C_{21}})) && (4.39)\\
x_{21} &= -C_{21}(1 - e^{-a_{21}x_{21}/C_{21}}) + \alpha_{21}(t)(\lambda_2(t) + C_{12}(1 - e^{-a_{12}x_{12}/C_{12}})) && (4.40)\\
x_{23} &= -C_{23}(1 - e^{-\frac{a_{23}x_{23}}{C_{23}}}) + \alpha_{23}(\lambda_2(t) + C_{12}(1 - e^{-a_{12}x_{12}/C_{12}})) && (4.41)
\end{aligned}
$$

Linearised around the origin (i.e $x_{jk} = 0$), the flow out function as given by the above model equals that of the model considered earlier in Section 4 2 In other words, for small values of the buffer occupancies the value of the flow out function as given by the exponential model and the linear model of the previous sections are nearly the same Similarly for large values of the buffer occupancies, the flowout is equal to the channel capacity of the link in both the cases For this class of models, consider an optimal routing strategy that minimises the total buffer occupancy time

defined earlier as

$$J = \int_0^T (x_{12} + x_{13} + x_{21} + x_{23})dt$$

It can be argued that the optimal routing strategy is a *bang-bang* type of control with following values for the routing variables

- If $p_{13}(t) < p_{12}(t)$, then $\alpha_{13}(t) = 1$
- If $p_{13}(t) > p_{12}(t)$, then $\alpha_{13}(t) = 0$.
- If $p_{23}(t) < p_{21}(t)$, then $\alpha_{23}(t) = 1$.
- If $p_{23}(t) > p_{21}(t)$, then $\alpha_{23}(t) = 0$

The differential equations governing the dynamics of the costate variables are

$$p_{13} = -1 + a_{13}e^{-a_{13}x_{13}/C_{13}}p_{13} \tag{4.42}$$

$$p_{12} = -1 + a_{12}e^{-a_{12}x_{12}/C_{12}}(p_{12} - \alpha_{21}p_{21} - \alpha_{23}p_{23}) \tag{4.43}$$

$$\dot{p}_{21} = -1 + a_{21}e^{-a_{21}x_{21}/C_{21}}(p_{21} - \alpha_{12}p_{12} - \alpha_{13}p_{13}) \tag{4.44}$$

$$p_{23} = -1 + a_{23}e^{-a_{23}x_{23}/C_{23}}p_{23} \tag{4.45}$$

with the transversality conditions $p_{12}(T) = p_{13}(T) = p_{21}(T) = p_{23}(T) = 0$.

The two-point boundary value problem in the state and costate variables was numerically integrated for various values of the link parameters and for various initial conditions as given in the Table 4.6. This numerical experiment was repeated for various input load patterns given in Table 4.7

From the above numerical investigations (examples of which are shown in the form of graphs as in Figure 4.31, Figure 4.32 and Figure 4.33), the following observations can be made regarding the nature of the optimal routing strategy

1. The optimal routing strategy has the **loop-free** property
2. There is at most one inter regime switching

Table 4 6· Case 1: Load $\lambda_1(t) \equiv \lambda_2(t) \equiv 5$

No	Link parameters	Initial Buffer Occupancies	Optimal Strategy
a	$a_{12} = a_{21} = 0\ 1$, $a_{13} = a_{23} = 0\ 5$, $C_{12} = C_{21} = C_{23} = 100000$ $C_{13} = 10$	$x_{12} = x_{13} = 0$ $x_{21} = x_{23} = 0$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$
b	same as above	$x_{12} = x_{21} = 10$ $x_{13} = 50, x_{23} = 10$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$
c	$a_{12} = a_{21} = 0\ 1$, $a_{13} = a_{23} = 0\ 5$, $C_{12} = C_{21} = 100000$ $C_{13} = 10, C_{23} = 15\ 0$	$x_{12} = x_{13} = 0$ $x_{21} = x_{23} = 0$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$
d	same as above	$x_{12} = x_{21} = 10$ $x_{13} = x_{23} = 50$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$
e	$a_{12} = a_{23} = 0\ 5$ $a_{13} = a_{21} = 0\ 1$ $C_{12} = C_{21} = C_{13} = 100000$ $C_{23} = 10$	$x_{12} = x_{21} = 0$ $x_{13} = x_{23} = 0$	$\alpha_{13}(t) = 0$, for $t \in [0, 3\ 6]$ $\alpha_{13}(t) = 1$, for $t \in (3\ 6, 10]$. $\alpha_{23}(t) \equiv 1$.
f	same as in 5	$x_{12} = x_{13} = x_{21} = 10$ $x_{23} = 50$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$.
g	same as in 5 except that $C_{13} = 10$	$x_{12} = x_{13} = 0$ $x_{21} = x_{23} = 0$	$\alpha_{13}(t) = 0$, for $t \in [0, 3\ 9]$ $\alpha_{13}(t) = 1$, for $t \in (3\ 9, 10]$. $\alpha_{23} \equiv 1$
h	same as in 7	$x_{12} = x_{13} = x_{21} = 10$ $x_{23} = 50$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$.
i	$a_{12} = a_{23} = 0\ 1$ $a_{21} = 0\ 6, a_{13} = 0\ 5$ $C_{12} = C_{21} = 100000$ $C_{13} = 8\ 0, C_{23} = 10$	$x_{12} = x_{13} = 0$. $x_{21} = x_{23} = 0$	$\alpha_{13}(t) \equiv 1$. $\alpha_{21}(t) = 1$ for $t \in [0, 3\ 0]$ $\alpha_{21}(t) = 0$, for $t \in (3\ 0, 10]$
j	same as in 9	$x_{12} = x_{21} = 5$ $x_{23} = 5, x_{23} = 30$	$\alpha_{13}(t) \equiv 1$ $\alpha_{23}(t) \equiv 1$

3 The system always ends in regime 4 [2] where direct paths are used at both the source nodes.

[2]The definitions of the regimes are the same as in Section 4 2

Table 4 7 Load pattern for the various cases.

Case	Load λ_1	Load λ_2
2	10	10
3	15	15
4	20	20
5	25	25
6	5	25
7	10	20
8	25	5
9	20	10
10	15	20

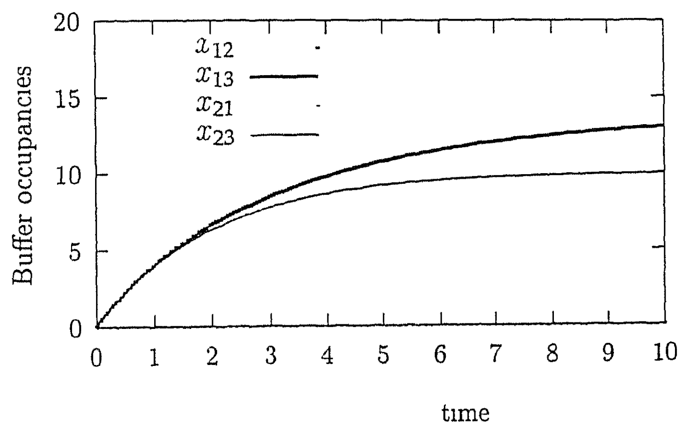

Figure 4 31: Buffer occupancies for the case 1a. The network operates in regime 4 for the entire duration.

Comments:

It was also observed that the numerical algorithm (which uses a Runge-Kutta-Merson method and a Newton iteration in a shooting and matching technique [24]) used for solving the two point boundary value problem in the state and costate variables is sensitive to the choice of the initial value of the costate variables, particularly when the link capacities are small (the links of the network operate in near saturation mode). Convergence was found to be difficult to obtain under this situation

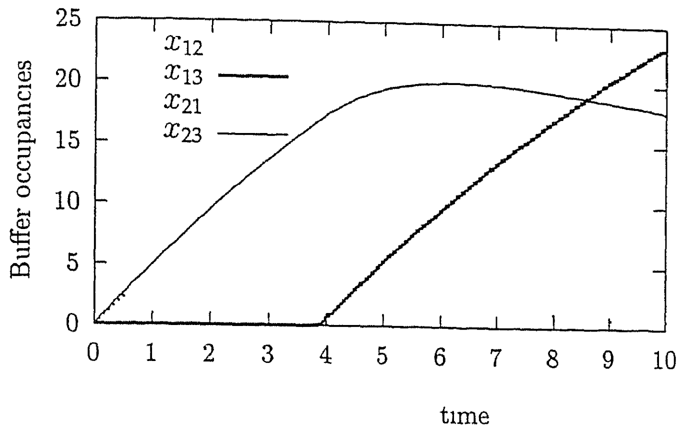

Figure 4.32: Buffer occupancies for case 1g The network operation switches from regime 2 to regime 4 at $t_s = 3\ 9$.

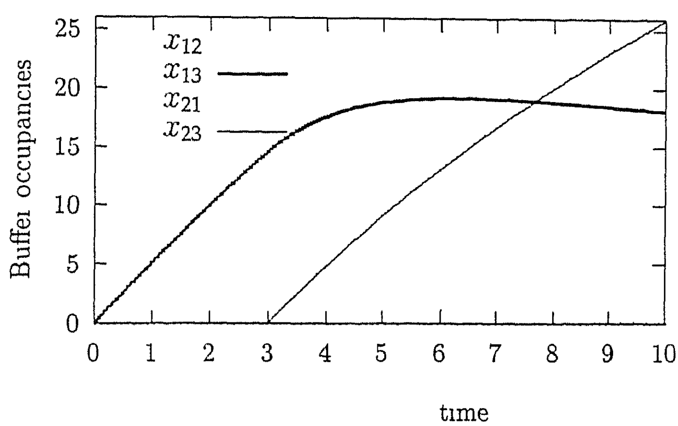

Figure 4 33 Buffer occupancies for case 1i. The network operation switches from regime 3 to regime 4 at $t_s = 3\ 0$.

The above numerical investigation raises certain interesting possibilities If the optimal routing strategy is characterised by the single-switching property with the terminal regime always being regime 4, then this property can be exploited in specifying the routing strategy in the following manner. If we know the nature of switching (i.e whether a regime 2 to regime 4 or a regime 3 to regime 4), then consider the performance of a suboptimal strategy in which this switching instant is an arbitrary value in $[0, T]$. By varying this switching instant between 0 to T, and

finding the instant at which the performance index reaches the minimum (which can be done graphically), the switching instant corresponding to the optimal strategy can be obtained The obvious advantage of this procedure is that it eliminates the need for solving the tedious two point boundary value problem in the state and costate variables Whether this property of *single switching* can be analytically proved and whether a necessary and sufficient condition which the link parameters and the input traffic have to satisfy for a regime 2 to regime 4 (or regime 3 to regime 4) switching to take place can be obtained, have not been investigated in this thesis

4.7 Conclusions

We investigated the problem of optimal traffic routing in a simple network unit of three nodes in this chapter. As in the Chapter 3 of this thesis, our emphasis was on the model in which the flow out function (for any link) depends linearly on the associated buffer occupancy and saturates at a value equal to the channel capacity The optimal routing strategy was synthesised for the case where in all the links of the network have infinite channel capacities Under this situation it was shown that the optimal strategy depends only on the topological parameters of the network and is independent of the input traffic. This in turn allows for an on-line implementation scheme We also argued that the optimal routing strategy is bang-bang in nature and has at the most one inter regime switching instant The optimal routing strategy was shown to have the **loop-free** property also.

We then relaxed the assumption that all links have infinite channel capacities and considered the case wherein the direct link (1,3) is of finite channel capacity. The **loop-free** property was shown to hold good under this situation also However the routing strategy need not be *bang-bang* and there could be *intervals of partial routing* (as was shown in the examples cited) The network operation was shown to end always in one of the four modes **FES**, **PET**, **EET** and **EEL**

We then considered networks (with finite C_{13}) in which the link parameters satisfy certain assumptions (as given by the conditions stated in Theorem 4.3 2)

The implication of these assumptions being that if all the links were of infinite channel capacities, then the optimal strategy would be a direct routing for the entire duration of network operation Under the additional assumption that the initial buffer occupancy on link (1,3) is below the saturation value, we derived a set of equations (in terms of the link parameters and the input traffic), whose solution specify the optimal strategy Since these equations are analytically difficult to solve and furthermore, need the knowledge of the input traffic for the entire duration of operation (thus an on-line implementation scheme is not possible), a suboptimal algorithm was proposed. The performances of the optimal and the suboptimal algorithm were compared in the case of some specific load patterns

Finally we investigated the nature of the optimal routing strategy in the case of an exponential model From the numerical investigation conducted, we observed that the optimal routing strategy has the **loop-free** property and the single switching property as in the case of the linear model Furthermore, it was also observed that the network operation always ends in regime 4. Whether or not these properties can be analytically proved and whether they can be exploited in specifying the optimal routing strategy by means of a simple procedure, are open issues which have not been addressed to in this thesis

Chapter 5

Optimal and Suboptimal Routing Strategies in Larger Networks

5.1 Introduction

The previous chapters have addressed to the issue of optimal/suboptimal traffic routing in two simple networks units, namely a two-node network in Chapter 3 and a three-node network in Chapter 4 As mentioned in Chapter 1, large (communication) networks can be viewed as being composed of simpler network units and therefore the problem of synthesising optimal (or at least good suboptimal) routing strategies for such large networks can be approached by considering routing strategies which are *locally* optimal (or at least good suboptimal) for the network units which compose them. With this perspective we now investigate the problem of optimal routing in some network topologies which are composed of simpler units analysed earlier. An overview of what is attempted in this chapter is as follows·

In Section 5 2, we study the nature of optimal routing strategy in a communication network between a source and a destination point in which there is a set of two parallel links between every successive pair of nodes. External traffic arriving at the source node and the transit nodes are to be routed to the destination via the subsequent network units. This topology essentially is a concatenation of the

two-node network units considered in the Chapter 3. We first consider the general case wherein m such network units (of the total n which constitute the topology) have a link (the *faster* link of the two) of finite channel capacity (both the links of the remaining $(n-m)$ units are assumed to be of infinite channel capacity) We show that in all the units which have both the links of infinite channel capacity, all the incoming traffic is routed onto the *faster* link for the entire duration of network operation For those units which have a link of finite channel capacity, either all the incoming traffic is routed onto the *faster* link or there is a partial routing on this link During intervals of such *partial routing*, a fraction of the traffic equal to the ratio of the channel capacity to the total incoming traffic is routed onto this link The properties of the optimal routing strategy which we proved in Chapter 3 (as given by Theorems 3.3 1, 3 3 2 and 3 3 3) when these units were considered in isolation are shown to hold good even when they are considered in conjunction with other units For the case where only one unit has a link of finite channel capacity (corresponding to a choice of m equal to 1), we obtain a set of equations (in terms of the link parameters and the input traffic to the network), which are to be solved to specify the optimal routing strategy. The specification of the optimal strategy requires the knowledge of the load patterns for the entire duration of network operation. Since this necessitates an off-line computation, an on-line implementable suboptimal algorithm is suggested along the lines as done for the individual network unit in Chapter 3. Examples are given wherein the performance of the optimal and suboptimal routing strategies are compared.

In Section 5.3, we look at a topology composed of the three-node structure of Chapter 4. We had, in Chapter 4, synthesised the optimal routing strategy for the case wherein all the links of this network unit have infinite channel capacity. When all the links of this topology are of infinite channel capacity, the performances of the routing strategy which is *globally* optimal (in the sense of minimising the total buffer occupancy time for the entire network) and the routing strategy synthesised from *locally* optimal ones are numerically compared for various choice of link parameters. When one of the direct links of one unit is of finite channel capacity, two

schemes of operating the network under the linear mode are considered The first one in which the linear optimal strategy (which is *globally* optimal) is implemented on the network and excess traffic which drives the link of finite channel capacity into saturation is rerouted onto a different unit in the topology. The second case corresponds to the implementation of *locally* optimal (linear) strategy with the excess traffic routed onto a different unit in the same layer of the topology The performance of these two schemes are compared for different link parameters of the network

Finally in Section 5 4, we consider a four-node hub network in which all the links are of infinite channel capacity. Topologically this network can be viewed as being composed of the three node network units which we considered in Chapter 4, wherein adjacent units share a common direct link to the destination It is shown that the optimal routing strategy for this network unit has the *loop-free* property (This property was established for the constituent three-node network unit in the Chapter 4). The network operation always ends with a direct routing of packets at all the three source nodes The conditions on the link parameters under which the optimal strategy for this network is the same as that synthesised from the *locally optimal* strategies for the constituent units, is arrived at When these conditions are violated, a suboptimal way of traffic routing in this network can be obtained from the *locally optimal* strategies for the network units. We compare the performance of this algorithm with that of the optimal one in the case of some numerical examples

5.2 Network topology of the two node network units

Consider a network topology shown in Figure 5 1, in which the units k_i, $i = 1, 2, \ldots m$ have a link of finite channel capacity and all the remaining $(n - m)$ units are of links of infinite channel capacity The following assumptions are made in the analysis that follows·

Assumptions :

Assumption 5.1.1 The faster link in the k_i-th unit has a finite channel capacity

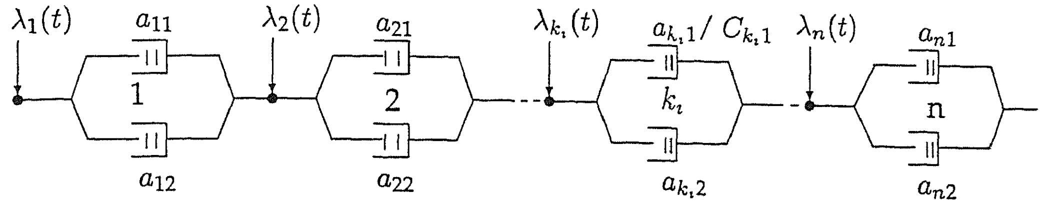

Figure 5.1 A network topology of two node units

equal to C_{k_i1}, $i = 1, 2, \quad m$

Assumption 5.1.2 The initial buffer occupancies for the links with finite channel capacities do not exceed their saturation values, i.e $x_{k_i1}(0) < x_{k_i1s}$ for $i = 1, 2, \quad m$

Assumption 5.1.3 All the buffers are of infinite capacity and therefore packets are never rejected from any link

As in the previous chapters the optimal routing strategy is defined as one which minimises the total buffer occupancy time for the network in the entire duration of network operation.

The system dynamics in this case is given by the following set of equations

$$
\begin{aligned}
x_{11} &= -a_{11}x_{11} + \alpha_{11}(t)\lambda_1(t) \\
x_{12} &= -a_{12}x_{12} + \alpha_{12}(t)\lambda_1(t) \\
x_{21} &= -a_{21}x_{21} + \alpha_{21}(t)(\lambda_2(t) + a_{11}x_{11} + a_{12}x_{12}) \\
x_{22} &= -a_{22}x_{22} + \alpha_{22}(t)(\lambda_2(t) + a_{11}x_{11} + a_{12}x_{12}) \\
. &= \vdots \\
x_{k_11} &= -f_{k_11}(x_{k_11}) + \alpha_{k_11}(t)(\lambda_{k_1}(t) + a_{(k_1-1)1}x_{(k_1-1)1} + a_{(k_1-1)2}x_{(k_1-1)2}) \\
x_{k_12} &= -a_{k_12}x_{k_12} + \alpha_{k_12}(t)(\lambda_{k_1}(t) + a_{(k_1-1)1}x_{(k_1-1)1} + a_{(k_1-1)2}x_{(k_1-1)2}) \\
. &= . \\
x_{k_m1} &= -f_{k_m1}(x_{k_m1}) + \alpha_{k_m1}(t)(\lambda_{k_m}(t) + a_{(k_m-1)1}x_{(k_m-1)1} + a_{(k_m-1)2}x_{(k_m-1)2}) \\
x_{k_m2} &= -a_{k_m2}x_{k_m2} + \alpha_{k_m2}(t)(\lambda_{k_m}(t) + a_{(k_m-1)1}x_{(k_m-1)1} + a_{(k_m-1)2}x_{(k_m-1)2}) \\
&= . \\
x_{n1} &= -a_{n1}x_{n1} + \alpha_{n1}(t)(\lambda_n(t) + a_{(n-1)1}x_{(n-1)1} + a_{(n-1)2}x_{(n-1)2})
\end{aligned}
$$

$$x_{n2} = -a_{n2}x_{n2} + \alpha_{n2}(t)(\lambda_n(t) + a_{(n-1)1}x_{(n-1)1} + a_{(n-1)2}x_{(n-1)2})$$

where

$$\begin{aligned} f_{k,1}(x_{k,1}) &= a_{k,1}x_{k,1} \quad \text{if } x_{k,1} < x_{k,1s} \\ &= C_{k,1} \quad \text{otherwise} \end{aligned}$$

Because of the discontinuity of the function $f_{k,1}(x_{k,1})$ at the point $x_{k,1} = x_{k,1s}$, the Maximum Principle is not directly applicable. We therefore approximate this function by $f^r_{k,1}(x_{k,1})$, which is obtained by drawing an arc of radius r tangential to the portions $a_{k,1}x_{k,1}$ and the constant $C_{k,1}$ of the original function $f_{k,1}(x_{k,1})$ Corresponding to a value r, let the optimal state variables be denoted by $x^r_{11}(t)$, $x^r_{12}(t)$, $x^r_{21}(t)$, $x^r_{22}(t)$, . , $x^r_{k_11}(t)$, $x^r_{k_12}(t)$, ,$x^r_{k_m1}(t)$, $x^r_{k_m2}(t)$,$x^r_{n1}(t)$ and $x^r_{n2}(t)$, the costate variables be denoted by $p^r_{11}(t)$, $p^r_{12}(t)$,$p^r_{21}(t)$, $p^r_{22}(t)$, ..., $p^r_{k_11}(t)$, $p^r_{k_12}(t)$,. $p^r_{k_m1}(t)$, $p^r_{k_m2}(t)$, .., $p^r_{n1}(t)$, and $p^r_{n2}(t)$, and the routing variables be denoted by $\alpha^r_{12}(t)$, $\alpha^r_{12}(t)$, ..., $\alpha^r_{k_11}(t)$, $\alpha^r_{k_12}(t)$, ..,$\alpha^r_{k_m1}(t)$, $\alpha^r_{k_m2}(t)$, . . , $\alpha^r_{n1}(t)$ and $\alpha^r_{n2}(t)$

Remark 5.2.1 The value of $x^r_{k,1}(t)$ at which the function $\frac{df^r_{k,1}}{dx^r_{k,1}}$ equals $a_{k,2}$ is denoted by $x_{k,tr}$

The performance index J to be minimised is the total buffer occupancy time

$$J = \int_0^T (x^r_{11}(t) + x^r_{12}(t) + \ . + x^r_{k,1}(t) + x^r_{k,2}(t) + \quad + x^r_{n1}(t) + x^r_{n2}(t))dt$$

The Hamiltonian is given as

$$\begin{aligned} H^r = \; & (x^r_{11} + x^r_{12} + \ . \ + x^r_{k_11} + x^r_{k_12} + \quad + x^r_{k_m1}(t) + x^r_{k_m2}(t) + \quad + x^r_{n1} + x^r_{n2}(t)) \\ & + p^r_{11}(-a_{11}x^r_{11} + \alpha^r_{11}(t)\lambda_1(t)) \\ & + p^r_{12}(-a_{12}x^r_{12} + \alpha^r_{12}\lambda_1) \\ & + \\ & + p^r_{k_11}(-f^r_{k_11}(x^r_{k_11}) + \alpha^r_{k_11}(\lambda_{k_1} + a_{(k_1-1)1}x^r_{(k_1-1)1} + a_{(k_1-1)2}x^r_{(k_1-1)2})) \\ & + p^r_{k_12}(-a_{k_12}x^r_{k_12} + \alpha^r_{k_12}(\lambda_{k_1} + a_{(k_1-1)1}x^r_{(k_1-1)1} + a_{(k_1-1)2}x^r_{(k_1-1)2})) \end{aligned}$$

$$
\begin{aligned}
&+ \ \cdot \\
&+p^r_{k_m1}(-f^r_{k_m1}(x^r_{k_m1}) + \alpha^r_{k_m1}(\lambda_{k_m} + a_{(k_m-1)1}x^r_{(k_m-1)1} + a_{(k_m-1)2}x^r_{(k_m-1)2})) \\
&+p^r_{k_m2}(-a_{k_m2}x^r_{k_m2} + \alpha^r_{k_m2}(\lambda_{k_m} + a_{(k_m-1)1}x^r_{(k_m-1)1} + a_{(k_m-1)2}x^r_{(k_m-1)2})) \\
&+ \ \cdot \\
&+p^r_{n1}(-a_{n1}x^r_{n1} + \alpha^r_{n1}(\lambda_n + a_{(n-1)1}x^r_{(n-1)1} + a_{(n-1)2}x^r_{(n-1)2})) \\
&+p^r_{n2}(-a_{n2}x^r_{n2} + \alpha^r_{n2}(\lambda_n + a_{(n-1)1}x^r_{(n-1)1} + a_{(n-1)2}x^r_{(n-1)2}))
\end{aligned}
$$

It is easily argued that the Hamiltonian minimisation with respect to the routing variables results in the following choice of the optimal routing variables
For $j = 1, 2, \quad k_1, k_2, \quad , k_m, \quad n,$

- If $p^r_{j1}(t) > p^r_{j2}(t)$ then $\alpha^r_{j1}(t) = 0$ and $\alpha^r_{j2}(t) = 1$.
- If $p^r_{j1}(t) < p^r_{j2}(t)$ then $\alpha^r_{j1}(t) = 1$ and $\alpha^r_{j2}(t) = 0$.
- If during an interval $p^r_{j1}(t) = p^r_{j2}(t)$ then the routing variables are not specified by the above conditions

The equation for the costate variables are given as follows

$$
\begin{aligned}
\dot{p}^r_{11} &= -1 + a_{11}(p^r_{11} - \alpha^r_{21}p^r_{21} - \alpha^r_{22}p^r_{22}) \\
\dot{p}^r_{12} &= -1 + a_{12}(p^r_{12} - \alpha^r_{21}p^r_{21} - \alpha^r_{22}p^r_{22}) \\
\vdots \quad & \quad \vdots \\
\dot{p}^r_{k_11} &= -1 + \frac{df^r_{k_11}(x^r_{k_11})}{dx^r_{k_11}}(p^r_{k_11} - \alpha^r_{(k_1+1)1}p^r_{(k_1+1)1} - \alpha^r_{(k_1+1)2}p^r_{(k_1+1)2}) \\
\dot{p}^r_{k_12} &= -1 + a_{k_12}(p^r_{k_12} - \alpha^r_{(k_1+1)1}p^r_{(k_1+1)1} - \alpha^r_{(k_1+1)2}p^r_{(k_1+1)2}) \\
\vdots \quad & \quad \vdots \\
\dot{p}^r_{k_m1} &= -1 + \frac{df^r_{k_m1}(x^r_{k_m1})}{dx^r_{k_m1}}(p^r_{k_m1} - \alpha^r_{(k_m+1)1}p^r_{(k_m+1)1} - \alpha^r_{(k_m+1)2}p^r_{(k_m+1)2} \\
\dot{p}^r_{k_m2} &= -1 + a_{k_m2}(p^r_{k_m2} - \alpha^r_{(k_m+1)1}p^r_{(k_m+1)1} - \alpha^r_{(k_m+1)2}p^r_{(k_m+1)2}) \\
\vdots \quad & \quad . \\
\dot{p}^r_{n1} &= -1 + a_{n1}p^r_{n1} \\
\dot{p}^r_{n2} &= -1 + a_{n2}p^r_{n2}
\end{aligned}
$$

with the transversality conditions $p^r_{jl}(T) = 0$; for $j = 1, \ldots, n$ and for $l = 1, 2$

5.2.1 Properties of the Optimal routing strategy

Certain interesting observations can be made regarding the nature of the optimal routing strategy for this network topology. These are stated in the following lemmas and theorems.

Lemma 5.2.1 *The costate variables $p^r_{ji}(t)$ (for $j = 1, 2, \ldots, n$ and $i = 1, 2$) are all non-negative in $[0, T]$*

Proof : It is easily verified that $p^r_{n1}(t) = \frac{1-e^{a_{n1}(t-T)}}{a_{n1}}$ and $p^r_{n2}(t) = \frac{1-e^{a_{n2}(t-T)}}{a_{n2}}$ are non-negative in the interval $[0, T]$.

$$
\begin{aligned}
\dot{p}^r_{(n-1)1} &= -1 + a_{(n-1)1}(p^r_{(n-1)1} - \alpha^r_{n1}p^r_{n1} - \alpha^r_{n2}p^r_{n2}) \\
&\leq -1 + a_{(n-1)1}p^r_{(n-1)1} \;\; \forall t \in [0, T] \text{ and } p^r_{(n-1)1}(T) = 0.
\end{aligned}
$$

From the above inequality and the transversality condition it can be argued by the following steps that $p^r_{(n-1)1}(t)$ is greater than zero in the interval $[0, T)$

Compare $p^r_{(n-1)1}(t)$ with the function $p(t)$ where $p(t)$ satisfies the differential equation

$$
\begin{aligned}
\dot{p}(t) &= -1 + a_{(n-1)1}p, \quad p(T) = 0. \\
\text{i.e. } p(t) &= \frac{1 - e^{a_{(n-1)1}(t-T)}}{a_{(n-1)1}}
\end{aligned}
$$

We have

$$
\begin{aligned}
\dot{p}^r_{(n-1)1} - \dot{p} &\leq a_{(n-1)1}(p^r_{(n-1)1} - p) \\
&= a_{(n-1)1}(p^r_{(n-1)1} - p) - h(t) \quad \text{where } h(t) \geq 0
\end{aligned}
$$

Since $p^r_{(n-1)1}(T) - p(T) = 0$,

$$
p^r_{(n-1)1}(t) - p(t) = -e^{a_{(n-1)1}(t-T)} \int_T^t h(\tau)e^{a_{(n-1)1}(T-\tau)}d\tau
$$

$$
\begin{aligned}
&= e^{a_{(n-1)1}(t-T)} \int_t^T h(\tau) e^{a_{(n-1)1}(T-\tau)} d\tau \\
&\geq 0 \quad \forall t \in [0, T) \\
\text{Therefore} \quad p^r_{(n-1)1}(t) &\geq \frac{1 - e^{a_{(n-1)1}(t-T)}}{a_{(n-1)1}}
\end{aligned}
$$

It is similarly argued that,

$$
\begin{aligned}
p^r_{(n-1)2}(t) &\leq -1 + a_{(n-1)2} p^r_{(n-1)2} \quad p^r_{(n-1)2}(T) = 0 \\
\text{Therefore} \quad p^r_{(n-1)2}(t) &\geq \frac{1 - e^{a_{(n-1)2}(t-T)}}{a_{(n-1)2}} > 0, \quad \forall t \in [0, T)
\end{aligned}
$$

Since

$$
p^r_{(n-2)1}(t) = -1 + a_{(n-2)1}(p^r_{(n-1)2} - \alpha^r_{(n-1)1} p^r_{(n-1)1} - \alpha^r_{(n-1)2} p^r_{(n-1)2}
$$

and since $p^r_{(n-1)1}(t)$ and $p^r_{(n-1)2}(t)$ as proved above are non-negative in $[0, T)$, we get the relationship

$$
\begin{aligned}
p^r_{(n-2)1}(t) &\leq -1 + a_{(n-2)1} p^r_{(n-2)1}(t) \\
\text{Hence} \quad p^r_{(n-2)1}(t) &\geq \frac{1 - e^{a_{(n-2)1}(t-T)}}{a_{(n-2)1}} \\
\text{Similarly} \quad p^r_{(n-2)2}(t) &\geq \frac{1 - e^{a_{(n-2)2}(t-T)}}{a_{(n-2)2}}
\end{aligned}
$$

Therefore both $p^r_{(n-2)1}(t)$ and $p^r_{(n-2)2}(t)$ are non-negative in $[0, T]$

The above arguments can be easily applied for all the units between k_m and n. A similar argument (but with the following modification) can be used to prove that $p^r_{k_m1}(t)$ is non-negative in $[0, T]$.

$$
\begin{aligned}
\dot{p}^r_{k_m1} &= -1 + \frac{df^r_{k_m1}(x^r_{k_m1})}{dx^r_{k_m1}}(p^r_{k_m1} - \alpha^r_{(k_m+1)1} p^r_{(k_m+1)1} - \alpha^r_{(k_m+1)2} p^r_{(k_m+1)2}) \\
&\leq -1 + \frac{df^r_{k_m1}(x^r_{k_m1})}{dx^r_{k_m1}} p^r_{k_m1}
\end{aligned}
$$

Compare $p^r_{k_m1}(t)$ with the function $f(t)$ which satisfies the differential equation

$$
\begin{aligned}
f(t) &= -1 + \frac{df_{k^r_{m1}}(x^r_{km1})}{dx_{k_{m1}}} f(t), \quad f(T) = 0 \\
\text{i e} \quad f(t) &= \int_t^T e^{-\int_0^\tau \frac{df^r_{k_m1}}{dx^r_{k_m1}} d\tau}
\end{aligned}
$$

We have

$$p^r_{k_{m1}} - f(t) \leq \frac{df^r_{k_{m1}}}{dx^r_{k_{m1}}}(p^r_{k_{m1}} - f(t))$$

Note that the function $\frac{df^r_{k_{m1}}}{dx^r_{k_{m1}}}$ is non-negative Let $\frac{df^r_{k_{m1}}}{dx^r_{k_{m1}}}$ be denoted by $a(t)$ and $(p^r_{k_{m1}}(t) - f(t))$ be denoted by $y(t)$. Then

$$\begin{aligned}
y(t) &\leq a(t)y(t), \quad \forall t \in [0,T] \\
y(t) &= a(t)y(t) - h(t), \quad \text{where } h(t) \geq 0. \\
y(t) &= y(T)e^{\int_T^t a(\tau)d\tau} - e^{\int_T^t a(\tau)d\tau}\Big(\int_T^t e^{-\int_T^\tau a(\gamma)d\gamma} h(\tau)d\tau\Big) \\
&\geq 0, \quad \forall t < T
\end{aligned}$$

Thus $p'_{k_{m1}}(t) \geq f(t)$, $\forall t \in [0,T)$.

$$\begin{aligned}
p^r_{k_{m1}}(t) &\geq \int_t^T e^{-\int_0^\tau \frac{df^r_{k_{m1}}}{dx^r_{k_{m1}}} d\tau} \\
&> 0, \quad \forall t \in [0,T).
\end{aligned}$$

Thus the costate variable $p^r_{k_{m1}}(t)$ is non-negative in $[0,T]$

The above arguments can be easily extended to all the costate variables and hence the lemma.

□

Lemma 5.2.2 *In all the subunits of the network except the ones which have a link of finite channel capacity, all of the input traffic is routed onto the faster link for the entire duration of network operation*

Proof :

Without loss of generality, let us assume that the upper links of the subunit have larger link parameters (a_{ji}) than the corresponding one for the lower links. i e. $a_{11} > a_{12}$, $a_{21} > a_{22}$, $a_{k_i1} > a_{k_i2}$,... and $a_{n1} > a_{n2}$.

It suffices to show that $p^r_{11}(t) < p^r_{12}(t)$, $p^r_{21}(t) < p^r_{22}(t)$,. $p^r_{n1}(t) < p^r_{n2}(t)$, $\forall t \in [0.T)$ to prove the lemma

Since $a_{n1} > a_{n2}$,

$$p^r_{n1}(t) = \frac{1 - e^{a_{n1}(t-T)}}{a_{n1}} < p^r_{n2}(t) = \frac{1 - e^{a_{n2}(t-T)}}{a_{n2}}$$

Therefore $\alpha^r_{n1}(t) = 1, \forall t \in [0, T]$

$$\begin{aligned}\text{Since}\quad \dot{p}^r_{j1} &= -1 + a_{j1}(p^r_{j1} - \alpha_{(j+1)1}p^r_{(j+1)} - \alpha^r_{(j+1)2}p^r_{(j+1)2}) \\ \dot{p}^r_{j2} &= -1 + a_{j2}(p^r_{j2} - \alpha^r_{(j+1)1}p^r_{(j+1)1} - \alpha^r_{(j+1)2}p^r_{(j+1)2}) \qquad \text{for } j \neq k_1, k_2, \quad k_m\end{aligned}$$

We have

$$\begin{aligned}\dot{p}_{j2} - \dot{p}_{j1} &= (a_{j2}p^r_{j2} - a_{j1}p^r_{j1}) + (a_{j1} - a_{j2})(\alpha^r_{(j+1)1}p^r_{(j+1)1} + \alpha^r_{(j+1)2}p^r_{(j+1)2}) \\ &< a_{j2}p^r_{j2} - a_{j1}p^r_{j1} \quad \text{(since the second term above is non-negative).} \\ \text{Thus}\quad \dot{p}_{j2} - \dot{p}_{j1} &< (a_{j2}p^r_{j2} - a_{j1}p^r_{j1}) \\ &< a_{j2}(p^r_{j2} - p^r_{j1})\end{aligned}$$

We have the transversality condition $(p^r_{j2}(T) - p^r_{j1}(T)) = 0$ It can be argued that if the function $y(t)$ is such that $\dot{y} < ay$ (where $a > 0$) and $y(T) = 0$ then $y(t) > 0$, $\forall t < T$ Hence it follows that $p^r_{j2}(t) > p^r_{j1}(t), \forall t \in [0, T)$ and consequently $\alpha^r_{j1}(t) = 1$, $\forall t \in [0, T]$.

□

Thus the only network units in which a *partial routing* possibly take place are the ones with a link of finite channel capacity. We had earlier established in Chapter 3 in Lemma 3 3 1, that during an interval of *partial routing* the fraction of the traffic routed onto the *faster* link is the ratio of the channel capacity of the link to the total input traffic It is easily argued along the following lines, that in this network topology also, during any *interval of partial routing,* the fraction of traffic routed onto the faster link is the ratio of the channel capacity of the link to the total input traffic arriving at the network unit ·

During an interval in which the costate variables $p^r_{k_i1}(t)$ and $p^r_{k_i2}(t)$ are identical, their derivatives are also identical This, in turn, implies that the function $\frac{df^r_{k_i1}}{dx^r_{k_i1}}$ is

identically equal to $a_{k,2}$ during this interval and hence the buffer occupancy $x^r_{k,1}(t)$ is identically equal to $x_{k,tr}$. For the limiting case where r tends to zero, during such an interval in which $p_{k,1}(t)$ and $p_{k,2}(t)$ are identical, the buffer occupancy in link $k,1$ is identically equal to $x_{k,1s}$ and the routing variable $\alpha_{k,1}(t)$ is given by

$$\alpha_{k,1}(t) = \frac{C_{k,1}}{\lambda_{k_i}(t) + a_{(k_i-1)1}x_{(k_i-1)1} + a_{(k_i-1)2}x_{(k_i-1)2}}$$

It can be argued that the properties of the optimal routing strategy which we proved in Theorems 3.3.1, 3.3 2 and 3.3.3 when the network units k_i (i=1,2,. ... .m) are considered in isolation hold true when they are considered in conjunction with the other units as in the topology shown in Figure 5.1 We prove these results in two steps First we prove that the properties of the optimal routing strategy hold true for the k_m-th unit Subsequently we extend them to all other units (k_i, where i=1,2, .. m-1)

Lemma 5.2.3 *Let $a_{j1} > a_{j2}$, for $j = k_m, k_m+1, \quad, n$. Then the costate variables satisfy the following relationship*

$$p^r_{j1}(t) \quad > \quad p^r_{(j+1)1}(t), \text{for } j = k_m, \quad, (n-1) \text{ and } \forall t \in [0,T)$$

Proof: (By induction)
By Lemma 5 2.2, $\alpha^r_{j1}(t) = 1, \forall t \in [0,T]$ and for $j = k_m+1, k_m+2, \qquad n$. Therefore the differential equations for the costate variables $p^r_{j1}(t)$ and $p^r_{(j+1)1}(t)$ ($k_m < j < (n-1)$) are given as the following:

$$\begin{aligned} \dot{p}^r_{j1}(t) &= -1 + a_{j1}(p^r_{j1}(t) - p^r_{(j+1)1}(t)) \\ \dot{p}^r_{(j+1)1}(t) &= -1 + a_{(j+1)1}(p^r_{(j+1)1}(t) - p^r_{(j+2)1}(t)) \end{aligned}$$

Assume that the inequality stated in the lemma holds true for all $j \geq l$, for some l where $k_m < l < n$.
Then

$$\begin{aligned} \dot{p}^r_{(l-1)1}(t) &= -1 + a_{(l-1)1}(p^r_{(l-1)1}(t) - p^r_{l1}(t)) \\ \dot{p}^r_{l1}(t) &= -1 + a_{l1}(p^r_{l1}(t) - p^r_{(l+1)1}(t)) \\ \dot{p}^r_{(l-1)1}(t) - \dot{p}^r_{l1}(t) &< a_{(l-1)1}(p^r_{(l-1)1} - p^r_{l1}) \end{aligned}$$

From the transversality condition $p^r_{(l-1)1}(T) - p^r_{l1}(T) = 0$, and from the above inequality, it can be argued that $p^r_{(l-1)1}(t) > p^r_{l1}(t)$, for all t in $[0,T)$ Thus if the stated inequality holds true for all $j \geq l$, then it holds true for $j = (l-1)$ also For the case $l = (n-1)$, the function $p^r_{(n-1)1}(t)$ as given by the following expression $p^r_{(n-1)1}(t)$,

$$p^r_{(n-1)1}(t) = \frac{1 - e^{a_{(n-1)1}(t-T)}}{a_{(n-1)1}} + \frac{1 - e^{a_{n1}(t-T)}}{a_{n1}} + \frac{e^{a_{(n-1)1}(t-T)} - e^{a_{n1}(t-T)}}{a_{(n-1)1} - a_{n1}}$$

is greater than $p^r_{n1}(t)$, $\forall t \in [0,T]$. Therefore, by induction, it follows that for $j = (k_m + 1), \quad (n-1)$, $p^r_{j1}(t) > p^r_{(j+1)1}(t)$

For the case $j = k_m$, the argument can be extended with the following modification

$$\begin{aligned}
p^{\prime}_{k_m1}(t) &= -1 + \frac{df^r_{k_m1}}{dx^r_{k_m1}}(p^r_{k_m1}(t) - p^r_{(k_m+1)1}(t)) \\
p^{\prime}_{(k_m+1)1}(t) &= -1 + a_{(k_m+1)1}(p^r_{(k_m+1)1}(t) - p^r_{(k_m+2)1}(t)) \\
&> -1 \quad \text{as proved above.} \\
p^{\prime}_{k_m1}(t) - p_{(k_m+1)1} &< \frac{df^r_{k_m1}}{dx^r_{k_m1}}(p^r_{k_m1}(t) - p^r_{(k_m+1)1}(t))
\end{aligned}$$

From the transversality condition $p^{\prime}_{(k_m)}(T) - p^{\prime}_{(k_m+1)1}(T) = 0$ and from the above inequality, it can be concluded that $p^r_{k_m1}(t) > p^r_{(k_m+1)1}(t)$, $\forall t \in [0,T)$

Thus for all $j \in \{k_m, \quad (n-1)\}$, $p^r_{j1}(t) > p^r_{(j+1)1}(t)$

Hence the lemma.

□

Remark 5.2.2 : In the discussion that follows, the terms *linear regime, transition regime* and *saturation regime* and *linear mode, transition mode* and *saturation mode* are used in the same sense as in the definitions in Chapter 3 For example, if for all t in an interval I, $x^r_{k_i1}(t)$ is less than $x^r_{(k_i1)l}$, then such an interval I is termed as a *linear regime* for unit k_i and the network unit k_i is said to be operating in *linear mode* during the interval I If during an interval I, $x^r_{k_i1}(t)$ lies between $x^r_{(k_i1)l}$ and $x^r_{(k_i1)s}$ then such an interval I is termed as a *transition regime* and the network unit k_i is said to be operating in *transition mode* during I Similarly, if during an interval I, $x^r_{k_i1}(t)$ is greater than $x^r_{(k_i1)s}$, then such an interval I is termed as a *saturation regime* and the network unit k_i is said to be operating in the *saturation mode* during I

Theorem 5.2.1 *The network operation in the k_m-th unit can not end in*

(i) a saturation regime

(ii) a transition regime in which $x^r_{k_m 1}(t) > x_{k_m tr}$

The proof of this is quite along the lines of that for Theorem 3 3.1 and therefore is not reproduced here

Theorem 5.2.2 *The routing variable $\alpha^r_{k_m 1}(t)$ is either 1 or*

$$\frac{f^r_{k_m 1}(x_{k_m tr})}{\lambda_{k_m}(t) + a_{(k_m-1)1} x^r_{(k_m-1)1} + a_{(k_m-1)2} x^r_{(k_m-1)2}}$$

for all t in $[0, T]$

Proof :

We prove that $p^r_{(k_m 1)}(t) \leq p^r_{(k_m 2)}(t)$ for all $t \in [0, T]$ as follows:

Let $p^r_{k_m 1}(t) > p^r_{k_m 2}(t), \forall t \in I$ where $I = [t_0, t_1]$

Then $\alpha^r_{k_m 1}(t) = 0$, over I, and $x^r_{k_m 1}(t)$ is a monotonically decreasing function in I. As in the proof of Theorem 3 3.1, it can be argued that $x^r_{k_m 1}(t) \leq x_{k_m tr}, \forall t \in I$, and therefore $\frac{df^r_{k_m 1}(x^r_{k_m 1})}{dx^r_{k_m 1}} \geq a_{k_m 2}$.

Thus

$$\begin{aligned} p^r_{k_m 1} &= -1 + \frac{df^r_{k_m 1}}{dx^r_{k_m 1}}(p^r_{k_m 1} - p^r_{(k_m+1)1}) \quad (\text{by Lemma 5 2 2, } \alpha^r_{(k_m+1)1}(t) \equiv 1) \\ p^r_{k_m 2} &= -1 + a_{k_m 2}(p^r_{k_m 2} - p^r_{(k_m+1)1}) \end{aligned}$$

By Lemma 5.2.3, the term $(p^r_{k_m 1} - p^r_{(k_m+1)1})$ is non-negative in $[0, T]$ Therefore

$$\begin{aligned} p^r_{k_m 1} &\geq -1 + a_{k_m 2}(p^r_{k_m 1} - p^r_{(k_m+1)1}) \\ p^r_{k_m 2} &= -1 + a_{k_m 2}(p^r_{k_m 2} - p^r_{(k_m+1)1}) \\ p^r_{k_m 1} - p^r_{k_m 2} &\geq a_{k_m 2}(p^r_{k_m 1} - p^r_{k_m 2}) \\ &> 0 \quad (\text{as per our assumption that } p^r_{k_m 1}(t) > p^r_{k_m 2}(t)) \end{aligned}$$

The above inequality implies that the difference between $p^r_{k_m 1}(t)$ and $p^r_{k_m 2}(t)$ will monotonically increase. Along the lines of Theorem 3 3 2, it can be argued that will

result in the violation of the the transversality condition $(p^r_{k_m1}(T) - p^r_{k_m2}(T)) = 0$ Therefore the assumption of the existence of an interval I during which $p^r_{k_m1}(t)$ is greater than $p^r_{k_m2}(t)$ is incorrect. Consequently $p^r_{k_m1}(t) \leq p^r_{k_m2}(t)$, $\forall t \in [0, T]$. We know that if during an interval $p^r_{k_m1}(t)$ is less than $p^r_{k_m2}(t)$ then $\alpha^r_{k_m1}(t)$ equals 1 and if $p^r_{k_m1}(t)$ equals $p^r_{k_m2}(t)$ then $\alpha^r_{k_m1}(t)$ is given by the expression

$$\frac{f^r_{k_m1}(x_{k_m tr})}{\lambda_{k_m}(t) + a_{(k_m-1)1}x^r_{(k_m-1)1} + a_{(k_m-1)2}x^r_{(k_m-1)2}}$$

Hence the theorem.

□

Theorem 5.2.3 *An interval of partial routing in the k_m-th unit can not be*

(a) *followed by a transition regime in which $x^r_{k_m1}(t) < x_{k_m tr}$*

(b) *preceded by a transition regime in which $x^r_{k_m1}(t) > x_{k_m tr}$*

The proof of this theorem is along similar lines as that for Theorem 3 3 3 and is therefore not reproduced here.

We can easily extend the Lemma 5.2 3 and the Theorems 5 2 1, 5.2.2 and 5 2.3 to the other network units also

Lemma 5.2.4 *Let $a_{j1} > a_{j2}$, for $j \in \{1, 2, \quad (k_m - 1)\}$ Then the costate variables satisfy the relationship $p^r_{j1}(t) > p^r_{(j+1)1}(t)$, $\forall t \in [0, T)$ and for $j \in \{1, 2, \ ., (k_m - 1)\}$*

Proof :

We have already established in Lemma 5 2.3 that $p^r_{k_m1}(t) > p^r_{(k_m+1)1}(t)$ By Theorem 5 2 2, $p^r_{k_m}(t) \leq p^r_{k_m2}(t)$, $\forall t \in [0, T]$ Therefore the differential equations for the costate variables $p^r_{k_m-1}(t)$ and $p^r_{k_m1}(t)$ are given as the following

$$\begin{aligned} p^r_{(k_m-1)1} &= -1 + a_{(k_m-1)1}(p^r_{(k_m-1)1}(t) - p^r_{k_m1}(t)) \\ p^r_{k_m1} &= -1 + \frac{df^r_{k_m1}}{dx^r_{k_m1}}(p^r_{k_m1}(t) - p^r_{(k_m+1)1}(t)) \end{aligned}$$

By Lemma 5.2 2, $(p^r_{k_m1}(t) - p^r_{(k_m+1)1}(t))$ is positive in $[0, T)$ Therefore

$$p^r_{(k_m-1)1}(t) - p^r_{k_m1}(t) \ < \ a_{(k_m-1)1}(p^r_{(k_m-1)1} - p^r_{k_m1}(t))$$

The above inequality along with the transversality condition $p^r_{(k_m-1)1}(T) - p^r_{k_m 1}(T) = 0$ results in the condition $p^r_{(k_m-1)1}(t) > p^r_{k_m 1}(t)$, $\forall t \in [0, T)$ The arguments based on induction that were used in the proof of Lemma 5 2 2 can be extended for $j \in \{k_{(m-1)}, \ldots, (k_m - 1)\}$

Using the fact $(p^r_{k_{(m-1)}1}(t) - p^r_{k_{(m-1)}+1}(t))$ is non- negative in $[0, T)$, Theorem 5 2.2 can be extended to the network unit $k_{(m-1)}$ also, and it can be argued that $p^r_{k_{(m-1)}1}(t) \leq p^r_{k_{(m-1)}2}(t)$, $\forall t \in [0, T)$ The same sequence of arguments which were used in establishing that $p^r_{j1}(t) > p^r_{(j+1)1}(t)$ for all $j \in \{k_{(m-1)}, k_{(m-1)} + 1, \ldots, (k_m - 1)\}$ can now be extended for $j \in \{k_{(m-2)}, \ldots, (k_{(m-1)} - 1)\}$ In other words, $p^r_{j1}(t) > p^r_{(j+1)1}(t)$, $\forall j \in \{k_{(m-2)}, (k_{(m-2)} + 1), \ldots, (k_{(m-1)} - 1)\}$ Using the fact that $p_{k_{(m-2)}1} > p_{k_{(m-2)}2}(t)$, it can be proved along the lines of Theorem 5 2 2 that $p_{k_{(m-2)}1} \leq p^r_{k_{(m-2)}2}(t)$, $\forall t \in [0, T)$ The sequence of arguments can be extended to all other units in an exactly similar manner and hence the lemma.

□

Theorem 5.2.4 *The network operation in the k_i-th unit ($i = 1, 2, \ldots, (m-1)$) can not end in*

(i) a saturation regime

(ii) a transition regime in which $x^r_{k_i 1}(t) > x_{k_i tr}$

The proof is exactly along the lines of that for Theorem 3 3 1, which makes use of the fact that if the network operation ends in an interval in which $x^r_{k_i}(t) > x_{k_i tr}$, then it should have started with an initial buffer occupancy value which is greater than $x_{k_i tr}$. But this is not the case as per the assumption that the network operation starts in the *linear mode.*

Theorem 5.2.5 *The routing variable $\alpha^r_{k_i 1}(t)$ is either 1 or*

$$\frac{f^r_{k_i 1}(x_{k_i tr})}{\lambda_{k_i}(t) + a_{(k_i-1)1} x^r_{(k_i-1)1} + a_{(k_i-1)2} x^r_{(k_i-1)2}}$$

for all $t \in [0, T]$.

The proof of this is exactly along the lines of Theorem 5 2 2 and is therefore not reproduced here

Theorem 5.2.6 *An interval of partial routing in the k_i-th unit ($i = 1, 2, \quad (m-1)$) can not be ·*

(a) *followed by a transition regime in which $x^r_{k_i1}(t) < x_{k_itr}$*

(b) *preceded by a transition regime in which $x^r_{k_i1}(t) > x_{k_itr}$*

Proof :
It can be argued along the lines of the proof for (a) in Theorem 3 2.3 that if an *interval of partial routing* is followed by a *transition regime* in which $x_{k_itr} < x_{k_itr}$, then the function $(p^l_{k_i1}(t) - p^r_{k_i2}(t))$ continues to be a positive and monotonically increasing function till $t = T$. This results in the violation of the transversality condition $(p^r_{k_i1}(T) - p^l_{k_i2}(T)) = 0$.
It can also be argued (as in the proof of (b) of Theorem 3 2.3) that a necessary condition for an *interval of partial routing* to be preceded by a *transition regime* in which $x^l_{k_i}(t) > x_{k_itr}$, is that the network operation starts with an initial buffer occupancy $x^l_{k_i1}(0) > x_{k_itr}$ Since as per the assumption that the network operation starts in the *linear mode* (for all the units), this necessary condition is not satisfied and hence an *interval of partial routing* can not be preceded by a *transition regime* in which $x^l_{k_i1}(t) > x_{k_itr}$.

□

Thus we see that the properties of the optimal routing strategy which were proved when the network units are considered in isolation hold true even when they are considered in conjunction with the other units This allows us to make the following conclusions regarding the optimal strategy for this topology

- In all the subunits which have both the links of infinite channel capacity, the *globally optimal* routing strategy is the same as the *locally optimal* routing strategy. This corresponds to a routing of all the incoming packets onto the *faster* link for the entire duration of network operation

- In all other subunits (which have the *faster* link of finite channel capacity) either the entire traffic is routed onto the *faster* link or there are *intervals of partial routing* during which a fraction equal to the ratio of the channel capacity to the incoming traffic, is routed onto the *faster* link. These *intervals of partial routing* can not be be preceded by a *transition regime* in which $x^r_{k_i1}(t) > x_{k_itr}$, and can not be followed by a *transition regime* in which $x^r_{k_i1}(t) < x_{k_itr}$

In the next subsection we explore a procedure to specify these *intervals of partial routing* for the case of a topology in which only one subunit has a link of finite channel capacity (corresponding to the case $m = 1$)

5.2.2 Single unit with a link of finite channel capacity

In the discussion that follows we consider the case where r tends to zero. Let us also assume without any loss of generality, that the upper link in each unit is faster than the lower link In other words, $a_{11} > a_{12}, a_{21} > a_{22}, \ldots, a_{n1} > a_{n2}$. Therefore by Lemma 5 2.2, in all the units which have both the links of infinite channel capacity, all the incoming traffic is routed onto the upper link for the entire duration of network operation.

Assume that the total input traffic $\lambda_{k_1}(t) + a_{(k_1-1)1}x_{(k_1-1)1} + a_{(k_1-1)2}x_{(k_1-1)2}$ arriving in the k_1-th unit has 'n' positive crossings above the value equal to C_{k_11}. From the above theorems, it can be concluded that *intervals of partial routing*, if they exist, are of the type $[t_{s_1}, t_1]$, $[t_{s_2}, t_2]$, .. where the instants t_{s1}, t_{s2}, . correspond to those at which $x_{k_11}(t)$ reaches the value x_{k_11s}. At t_i and t_i^* we have the following conditions

$$x_{k_11}(t_i) = x_{k_11}(t_i^*) = x_{k_11s}.$$

The unit k_1 operates in the *saturation mode* during $[t_i, t_i^*]$ with $\alpha_{k_1}(t) = 1$ Therefore

$$\dot{x}_{k_11} = -C_{k_11} + \lambda_{k_1} + a_{(k_1-1)1}x_{(k_1-1)1} + a_{(k_1-1)2}x_{(k_1-1)2}$$

Hence $x_{k_11}(t_i^*) - x_{k_11}(t_i) = \int_{t_i}^{t_i^*} (\lambda_{k_1}(t) + a_{(k_1-1)1}x_{(k_1-1)1}(t) + a_{(k_1-1)2}x_{(k_1-1)2} - C_{k_11})dt$

During the interval $[t_i^*, t_{s_i+1}]$, link $k_1 1$ has linear dynamics. The dynamics of $p_{k_1 1}(t)$ during this interval is given by

$$p_{k_1 1} = -1 + a_{k_1 1}(p_{k_1 1} - p_{(k_1+1)1})$$

At $t = t_{s_i+1}$, $p_{k_1 1}(t_{s_i+1}) = p_{k_1 2}(t_{s_i+1})$. Therefore

$$p_{k_1 1}(t_i^*) = p_{k_1 2}(t_{s_i+1})e^{a_{k_1 1}(t_i^* - t_{s_i+1})} - e^{a_{k_1 1}t_i^*}\int_{t_{s_i+1}}^{t_i^*} e^{-a_{k_1 1}\tau}(1 + a_{k_1 1}p_{(k_1+1)1}(\tau))d\tau$$

and

$$p_{k_1 1}(t_i) = (t_i^* - t_i) + p_{k_1 2}(t_{s_i+1})e^{a_{k_1 1}(t_i^* - t_{s_i+1})} - e^{a_{k_1 1}t_i^*}\int_{t_{s_i+1}}^{t_i^*} e^{-a_{k_1 1}\tau}(1 + a_{k_1 1}p_{(k_1+1)1}(\tau))d\tau$$

Since $p_{k_1 1}(t_i) = p_{k_1 2}(t_i)$, we get the following equations in t_{s_i}, t_i and t_i^*

For $i = 1, 2, \quad n$,

$$\begin{aligned} p_{k_1 2}(t_i) = {} & (t_i^* - t_i) + p_{k_1 2}(t_{s_i+1})e^{a_{k_1 1}(t_i^* - t_{s_i+1})} \\ & -e^{a_{k_1 1}t_i^*}\int_{t_{s_i+1}}^{t_i^*} e^{-a_{k_1 1}\tau}(1 + a_{k_1 1}p_{(k_1+1)1}(\tau))d\tau \end{aligned} \quad (5\ 4)$$

As mentioned earlier, the expression for $p_{k_1 2}(t)$ and $p_{(k_1+1)1}(t)$ are obtained from solving the differential equations in them which can be done recursively by starting from $p_{n1}(t)$, followed $p_{(n-1)1}(t)$, and so on.

Thus we observe that to specify the instants $t_1, t_i^*, t_{s2}, t_2, t_2^*, t_{s_3}, t_3, t_3^*, \ . . \ t_n, t_n^*$, the above sets of Equations (5 1), (5.3) and (5.4) need to be solved

Example : Consider the case where $n = 3$, $k_1 = 2$ as shown in the Figure 5.2.

The equations for $p_{31}(t)$ and $p_{22}(t)$ are as follows.

$$\begin{aligned} p_{31}(t) &= \frac{1 - e^{a_{31}(t-T)}}{a_{31}} \\ p_{22}(t) &= \frac{1 - e^{a_{22}(t-T)}}{a_{22}} + \frac{1 - e^{a_{31}(t-T)}}{a_{31}} + \frac{e^{a_{22}(t-T)} - e^{a_{31}(t-T)}}{(a_{22} - a_{31})} \end{aligned}$$

The equations to be solved to specify the instants $t_1, t_1^*, t_{s2}, t_2, t_2^*, \ldots \ t_{sn}, t_n, t_n^*$ are as follows:

$$\int_{t_i}^{t_i^*} (\lambda_2(t) + a_{11}x_{11}(t) + a_{12}x_{12}(t))dt = C_{21}(t_i^* - t_i) \quad (5.5)$$

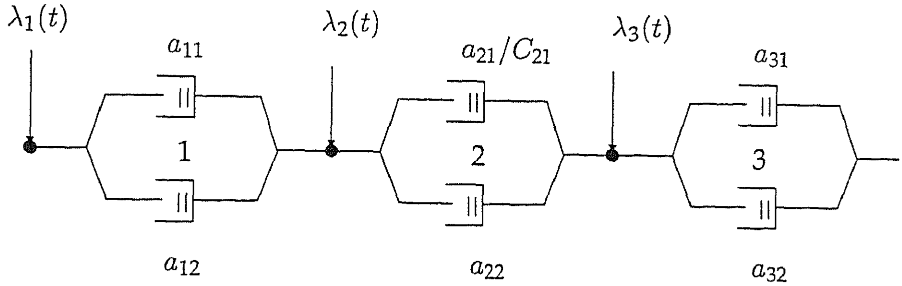

Figure 5 2 Network topology for $n = 3$, $k_1 = 2$

$$x_{21s}(1 - e^{-a_{21}(t_{s_i+1} - t_i^*)}) = e^{-a_{21}t_{s_i+1}} \int_{t_i^*}^{t_{s_i+1}} e^{a_{21}\tau}(\lambda_{21}(\tau) + a_{11}x_{11} + a_{12}x_{12})d\tau \tag{5.6}$$

$$\left.\begin{array}{l} \frac{1-e^{a_{22}(t_i - T)}}{a_{22}} + \frac{1-e^{a_{31}(t_i - T)}}{a_{31}} \\ + \frac{e^{a_{22}(t_i - T)} - e^{a_{31}(t_i - T)}}{(a_{22} - a_{31})} \end{array}\right\} = \left\{\begin{array}{l} (t_i^* - t_i) + [\frac{1-e^{a_{22}(t_{s_i+1} - T)}}{a_{22}} + \frac{1-e^{a_{31}(t_{s_i+1} - T)}}{a_{31}} + \\ \frac{e^{a_{22}(t_{s_i+1} - T)} - e^{a_{31}(t_{s_i+1} - T)}}{a_{22} - a_{31}}][e^{a_{21}(t_i^* - t_{s_i+1})}] - \\ (\frac{1-e^{a_{21}(t_i^* - t_{s_i+1})}}{a_{21}})(\frac{1}{a_{21}} + \frac{1}{a_{31}}) + \\ \frac{a_{21}e^{a_{31}(t_{s_i+1} - T)}(e^{a_{31}(t_i^* - t_{s_i+1})} - e^{a_{21}(t_i^* - t_{s_i+1})})}{a_{31}(a_{31} - a_{21})} \end{array}\right. \tag{5.7}$$

The function $x_{11}(t)$ in equation (5.4) is obtained by solving the differential equation

$$\begin{aligned} x_{11} &= -a_{11}x_{11} + \lambda_1(t) \\ \text{i e} \quad x_{11}(t) &= x_{11}(0)e^{-a_{11}t} + e^{-a_{11}t}\int_0^t \lambda_1(\tau)e^{a_{11}\tau}d\tau \\ \text{and} \quad x_{12}(t) &= x_{12}(0)e^{-a_{12}t} \end{aligned}$$

Thus we observe that for solving the above set of equations in t_1, t_1^*, t_{s2}, t_3, t_3^*, . .,t_{sn}, t_n, t_n^* we need to know the loads $\lambda_1(t)$ and $\lambda_2(t)$ for the entire duration $[0, T]$ Analytical solutions to these equations are difficult to obtain Therefore as suggested in the Chapter 3, consider the suboptimal algorithm in which we route the fraction of the incoming traffic needed to keep the link (2,1) flowout at its channel capacity ($\frac{C_{21}}{(\lambda_1(t)+a_{11}x_{11}+a_{12}x_{12})}$) till the input load to unit 2 ($\lambda_1(t)+a_{11}x_{11}+a_{12}x_{12}$) is greater than C_{21} The remaining traffic is routed onto link (2,2) during this interval Notice that this is also the suboptimal algorithm for the unit 2 when it is considered in isolation The examples given below give the relative performances

of the optimal strategy and the suboptimal strategy for some typical input load on the network. We also compare the performances of these strategies with the best performance that is achievable when all the links of the topology have infinite channel capacities.

5.2.3 Numerical Examples

Consider again the case where n =3, k=2 Let the link parameters be as follows:
$a_{11} = 0\,9$, $a_{12} = 0\,2$, $a_{21} = 0\,7$, $a_{22} = 0\,3$, $a_{31} = 0\,6$, $a_{32} = 0.1$, $C_{21} = 7$, $x_{21_s} = 10$
Let the network operation be for a duration $[0, 10]$

Example 5.2.3.1 :

Let $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = 5$ for the entire duration $[0, 10]$ Assume that the initial buffer occupancies are all zero.

The total traffic arriving at node 2 is given as.

$\lambda_2(t) + a_{11}x_{11} + a_{12}x_{12} = 5 + 5(1 - e^{-0\,9t})$, shown below:

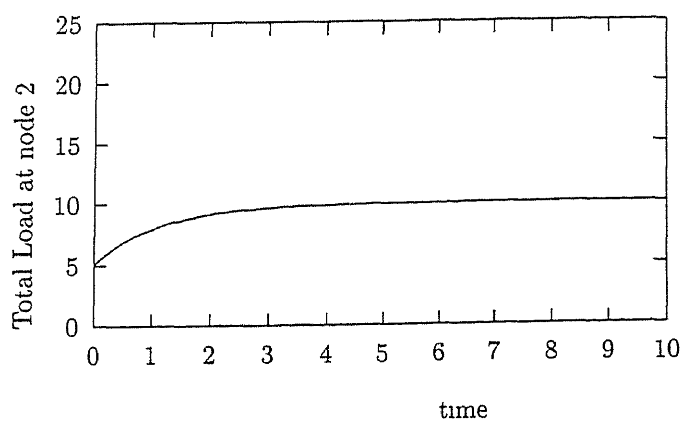

Figure 5 3. Total Load Arriving at Node 2 for Example 5.2.3.1

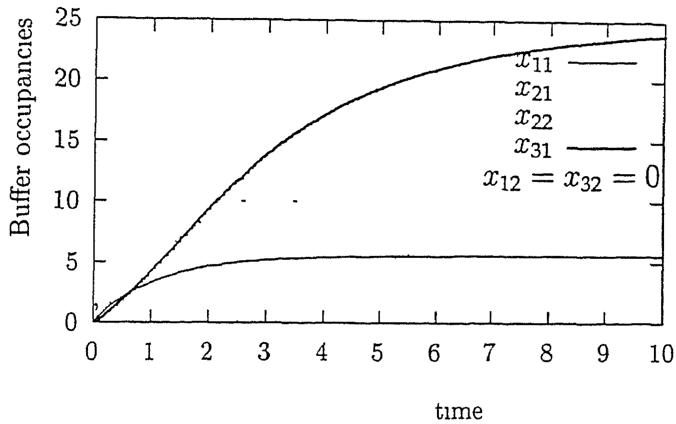

Figure 5 4 Buffer Occupancies under the Optimal and Suboptimal Strategies for Example 5 2.3 1

Table 5 1. Optimal and Suboptimal Strategies for Example 5 2.3 1

Optimal Strategy	Suboptimal Strategy
$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0\ 10]$	$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$
$\alpha_{21}(t) = 1$ over $[0, 2\ 4614]$ $\alpha_{21}(t) = \frac{7}{5+5(1-e^{-0\ 9t})}$ over $(2\ 4614, 10]$	$\alpha_{21}(t) = 1$ over $[0, 2\ 4614]$ $\alpha_{21}(t) = \frac{7}{5+5(1-e^{-0\ 9t})}$ over $(2\ 4614, 10]$
$J_{opt} = 341\ 7782$	$J_{subopt} = 341\ 7782$

Example 5.2.3.2: $\lambda_1(t) = \lambda_3(t) = 3\ 0$ for all $t \in [0\ 10]$

$$\lambda_2(t) = \begin{cases} 2\ , & \text{over } [0, 2\ 5] \\ 5\ , & \text{over } [2\ 5, 7\ 5] \\ 2\ ; & \text{over } [7\ 5, 10] \end{cases}$$

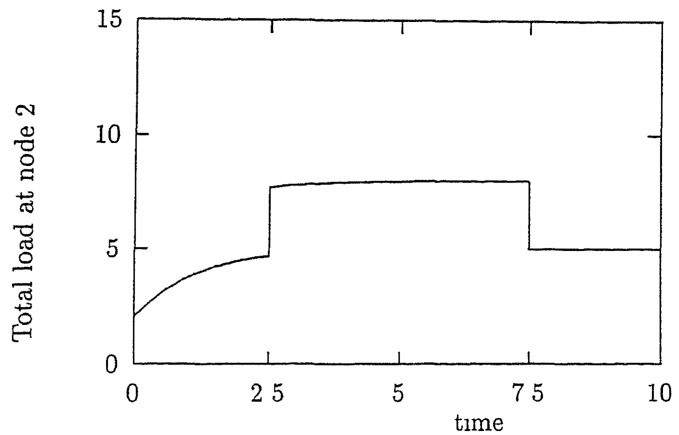

Figure 5.5: Total traffic arriving at node 2 for Example 5 2.3 2

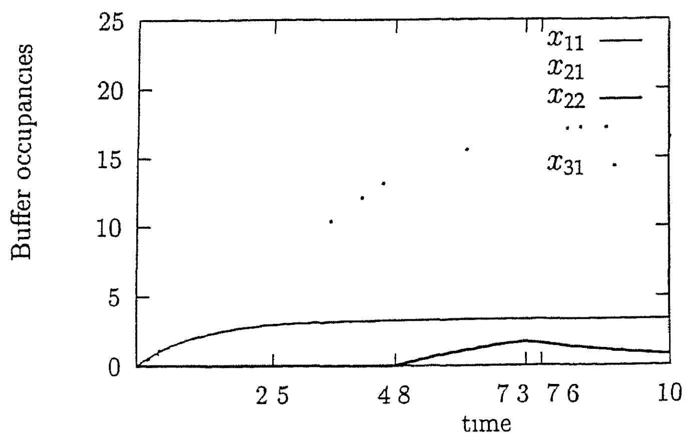

Figure 5 6 Buffer Occupancies under the Optimal Strategy for Example 5.2.3.2

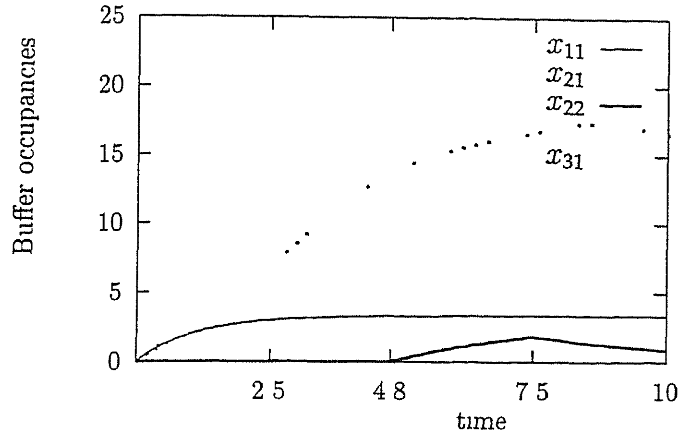

Figure 5 7 Buffer Occupancies under the Suboptimal Strategy for Example 5.2.3 2

Table 5 2: Optimal and Suboptimal Strategies for Example 5 2 3 2

Optimal Strategy	Suboptimal Strategy
$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$	$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$
$\alpha_{21}(t) = 1$ over $[0, 4\ 798]$ $\alpha_{21}(t) = \frac{7}{5+3(1-e^{-0\ 9t})}$ over $(4\ 798, 7\ 3]$	$\alpha_{21}(t) = 1$ over $[0, 4\ 798]$ $\alpha_{21}(t) = \frac{7}{5+3(1-e^{-0\ 9t})}$ over $(4\ 798, 7\ 5]$
$\alpha_{21}(t) = 1$ over $[7\ 3, 10]$	$\alpha_{21}(t) = 1$ over $[7\ 5, 10]$
$J_{opt} = 206\ 894$	$J_{subopt} = 208\ 322$

Example 5.2.3.3:

$$\lambda_1(t) = \begin{cases} 2 \ , & \text{over } [0, 2.5] \\ 5 \ ; & \text{over } [2\ 5, 7\ 5] \\ 2 \ , & \text{over } [7.5, 10] \end{cases}$$

$$\lambda_2(t) = \begin{cases} 2 \; ; & \text{over } [0, 2\ 5] \\ 5 \; ; & \text{over } [2\ 5, 7\ 5] \\ 2 \; ; & \text{over } [7\ 5, 10] \end{cases}$$

$$\lambda_3(t) = 5, \forall t \in [0, 10].$$

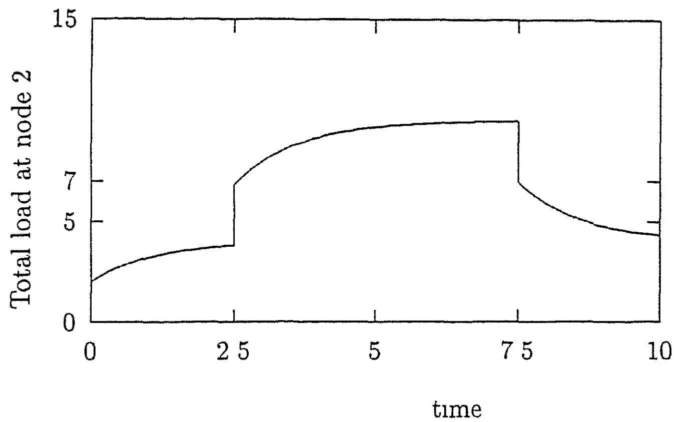

Figure 5 8. Total Load at Node 2 for Example 5 2 3 3

Table 5.3 Optimal and Suboptimal Strategies for Example 5 2.3.3

Optimal Strategy	Suboptimal Strategy
$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$	$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$
$\alpha_{21}(t) = 1$ over $[0, 4\ 3]$ $\alpha_{21}(t) = \frac{7}{10 - 3\ 2108e^{-0\ 9(t-2\ 5)})}$ over $(4\ 3, 7\ 25]$	$\alpha_{21}(t) = 1$ over $[0, 4\ 3]$ $\alpha_{21}(t) = \frac{7}{10 - 3\ 2108e^{-0\ 9(t-2\ 5)})}$ over $(4\ 3, 7\ 5]$
$\alpha_{21}(t) = 1$ over $[7\ 25, 10]$	$\alpha_{21}(t) = 1$ over $[7\ 5, 10]$
$J_{opt} = 286\ 68763$	$J_{subopt} = 296\ 78209$

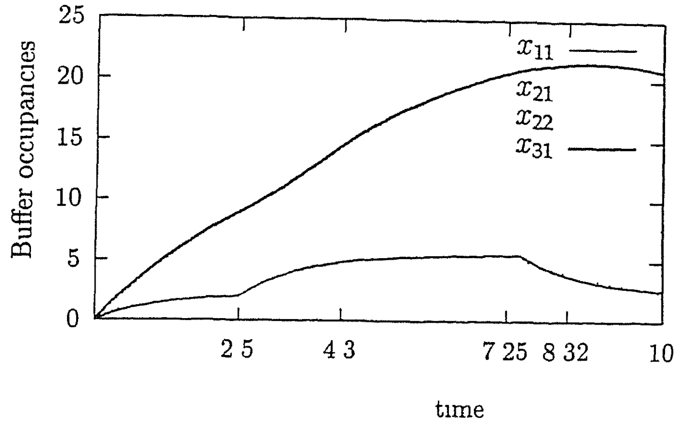

Figure 5 9 Buffer Occupancies under the Optimal Strategy for Example 5 2.3.3

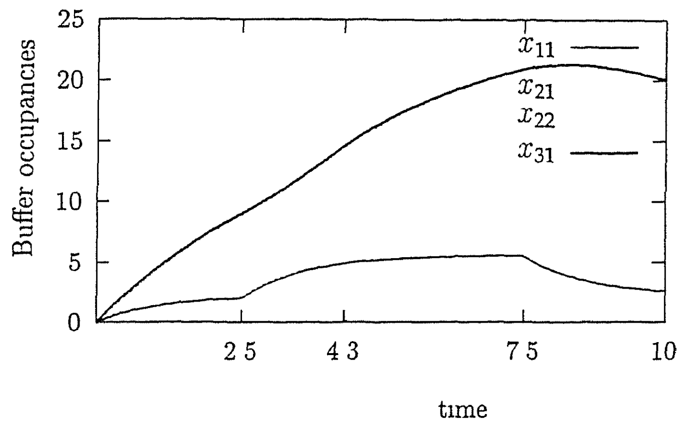

Figure 5.10 Buffer Occupancies under the Suboptimal Strategy for Example 5.2.3.3

Example 5.2.3.4:

$$\lambda_1(t) = \lambda_3(t) = 2, \forall t \in [0, 10].$$

$$\lambda_2(t) = \begin{cases} 2 & ; \text{ over } [0, 2.5] \\ 10 & ; \text{ over } [2.5, 5.0] \\ 2 & ; \text{ over } [5.0, 7.5] \\ 10 & ; \text{ over } [7.5, 10] \end{cases}$$

Table 5.4. Optimal and Suboptimal Strategies for Example 5.2.3.4

Optimal Strategy	Suboptimal Strategy
$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$	$\alpha_{11}(t) \equiv \alpha_{31}(t) \equiv 1, \forall t \in [0, 10]$
$\alpha_{21}(t) = 1$ over $[0, 3.386]$ $\alpha_{21}(t) = \frac{7}{12-2e^{-0.9t}}$ over $(3.386, 4.71]$	$\alpha_{21}(t) = 1$ over $[0, 3.386]$ $\alpha_{21}(t) = \frac{7}{12-2e^{-0.9t}}$ over $(3.386, 5]$
$\alpha_{21}(t) = 1$ over $[4.71, 8.04]$	$\alpha_{21}(t) = 1$ over $[5, 8.08]$
$\alpha_{21}(t) = \frac{7}{12-2e^{-0.9t}}$ over $(8.04, 10]$	$\alpha_{21}(t) = \frac{7}{12-2e^{-0.9t}}$ over $(8.08, 10]$
$J_{opt} = 220.92$	$J_{subopt} = 221.7122$

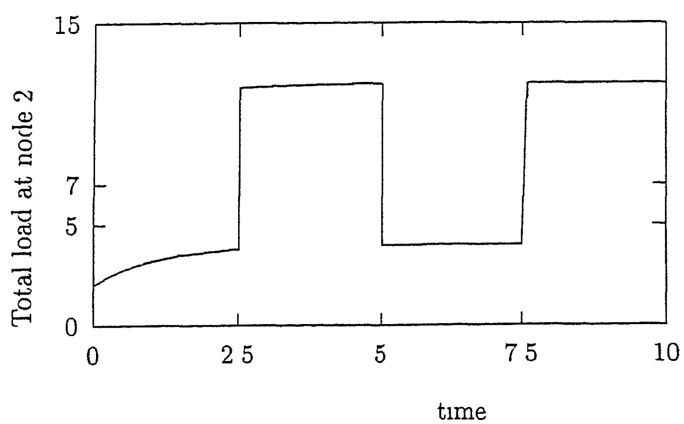

Figure 5.11 Total Load at Node 2 for Example 5.2.3.4

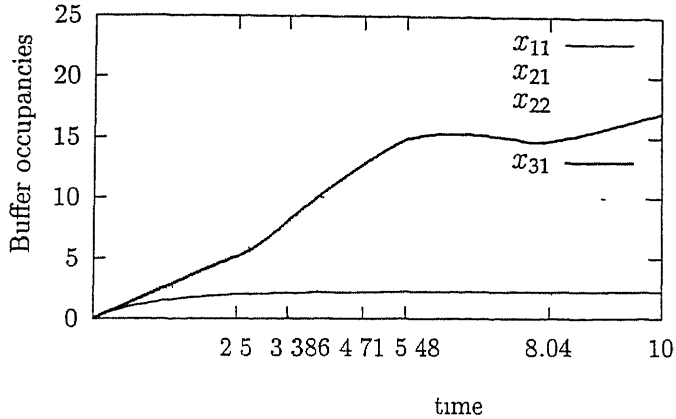

Figure 5 12: Buffer Occupancıes under the Optımal Strategy for Example 5 2 3 4

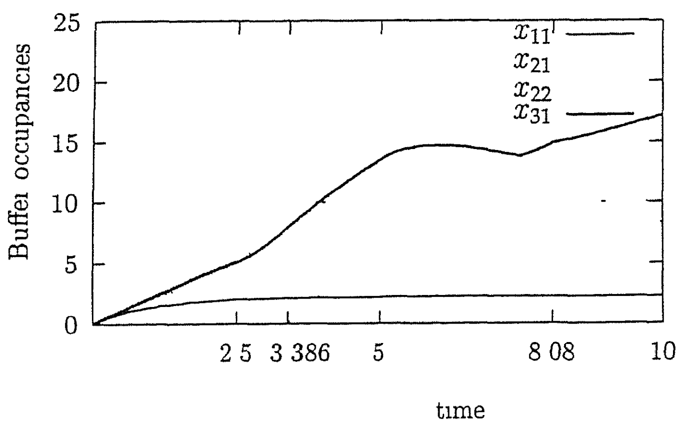

Fıgure 5 13 Buffer Occupancıes under the Suboptımal Strategy for Example 5 2.3.4

5.3 Network topology of three-node units.

Consider a network shown in Figure 5 14 composed of the three-node network units which we considered in Chapter 4 It was shown in Chapter 4, that if all the links of this network unit are of infinite channel capacity, then the routing strategy depends only on the link parameters of the network The optimal routing strategy was specified in terms of an inter-regime switching instant An equation (in the link parameters of the network and the total duration T of network operation was derived) the solution to which specifies this switching instant t_s

For the above network topology (shown in Figure 5 14), consider the *globally* optimal routing strategy which minimises the total buffer occupancy of the network

$$\int_0^T (x_{12} + x_{13} + x_{21} + x_{23} + x_{45} + x_{46} + x_{54} + x_{56} + x_{78} + x_{79} + x_{87} + x_{89})dt$$

for the case wherein all the units are of links of infinite channel capacity. The system dynamics for this case is given by

$$\dot{x}_{12} = -a_{12}x_{12} + \alpha_{12}(t)(\lambda_1(t) + a_{21}x_{21}) \quad (5\ 8)$$

$$x_{13} = -a_{13}x_{13} + \alpha_{13}(t)(\lambda_1(t) + a_{21}x_{21}) \quad (5.9)$$

$$x_{21} = -a_{21}x_{21} + \alpha_{21}(t)(\lambda_2(t) + a_{12}x_{12}) \quad (5.10)$$

$$\dot{x}_{23} = -a_{23}x_{23} + \alpha_{23}(t)(\lambda_2(t) + a_{12}x_{12}) \quad (5.11)$$

$$\dot{x}_{45} = -a_{45}x_{45} + \alpha_{45}(t)(\lambda_3(t) + a_{54}x_{54}) \quad (5.12)$$

$$x_{46} = -a_{46}x_{46} + \alpha_{46}(t)(\lambda_3(t) + a_{54}x_{54}) \quad (5.13)$$

$$\dot{x}_{54} = -a_{54}x_{54} + \alpha_{54}(t)(\lambda_4(t) + a_{45}x_{45}) \quad (5\ 14)$$

$$x_{56} = -a_{56}x_{56} + \alpha_{56}(t)(\lambda_4(t) + a_{45}x_{45}) \quad (5.15)$$

$$x_{78} = -a_{78}x_{78} + \alpha_{78}(t)(a_{13}x_{13} + a_{23}x_{23} + a_{87}x_{87}) \quad (5\ 16)$$

$$x_{79} = -a_{79}x_{79} + \alpha_{79}(t)(a_{13}x_{13} + a_{23}x_{23} + a_{87}x_{87}) \quad (5.17)$$

$$x_{87} = -a_{87}x_{87} + \alpha_{87}(t)(a_{46}x_{46} + a_{56}x_{56} + a_{78}x_{78}) \quad (5\ 18)$$

$$x_{89} = -a_{89}x_{89} + \alpha_{89}(t)(a_{46}x_{46} + a_{56}x_{56} + a_{78}x_{78}) \quad (5\ 19)$$

The Hamiltonian for this system H is given by

$$\begin{aligned} H &= x_{12} + x_{13} + x_{21} + x_{23} + x_{45} + x_{54} + x_{46} + x_{56} + x_{78} + x_{79} + \\ &\quad x_{87} + x_{89} + p_{12}\dot{x}_{12} + p_{13}x_{13} + p_{21}x_{21} + p_{23}x_{23} + p_{45}x_{45} + \\ &\quad p_{46}\dot{x}_{46} + p_{54}x_{54} + p_{56}\dot{x}_{56} + p_{78}\dot{x}_{78} + p_{79}\dot{x}_{79} + p_{87}\dot{x}_{87} + p_{89}x_{89} \end{aligned}$$

As in the case of the constituent three node unit, the optimal routing strategy for this composite network is *bang-bang* in nature with the following values for the routing variables.

- If during an interval I, $p_{jk}(t) > p_{jl}(t)$ then $\alpha_{jk}(t) = 0, \alpha_{jl}(t) = 1$

The costate variables $p_{jk}(t)$ are obtained by solving the following set of differential equations

$$p_{12} = -1 + a_{12}(p_{12} - \alpha_{21}p_{21} - \alpha_{23}p_{23}) \tag{5.20}$$

$$p_{13} = -1 + a_{13}(p_{13} - \alpha_{78}p_{78} - \alpha_{79}p_{79}) \tag{5.21}$$

$$p_{21} = -1 + a_{21}(p_{21} - \alpha_{12}p_{12} - \alpha_{13}p_{13}) \tag{5 22}$$

$$p_{23} = -1 + a_{23}(p_{23} - \alpha_{78}p_{78} - \alpha_{79}p_{79}) \tag{5.23}$$

$$p_{45} = -1 + a_{45}(p_{45} - \alpha_{54}p_{54} - \alpha_{56}p_{56}) \tag{5 24}$$

$$p_{54} = -1 + a_{54}(p_{54} - \alpha_{45}p_{45} - \alpha_{46}p_{46}) \tag{5 25}$$

$$p_{46} = -1 + a_{46}(p_{46} - \alpha_{87}p_{87} - \alpha_{89}p_{89}) \tag{5.26}$$

$$p_{56} = -1 + a_{56}(p_{56} - \alpha_{87}p_{87} - \alpha_{89}p_{89}) \tag{5 27}$$

$$\dot{p}_{78} = -1 + a_{78}(p_{78} - \alpha_{87}p_{87} - \alpha_{89}p_{89}) \tag{5.28}$$

$$p_{87} = -1 + a_{87}(p_{87} - \alpha_{78}p_{78} - \alpha_{79}p_{79}) \tag{5.29}$$

$$p_{79} = -1 + a_{79}p_{79} \tag{5 30}$$

$$p_{89} = -1 + a_{89}p_{89} \tag{5.31}$$

with the transversality conditions $p_{jk}(T) = 0$.

Since the set of equations in the state variables and the costate variables are decoupled, and since the routing variables depends only on the costate variables, it is

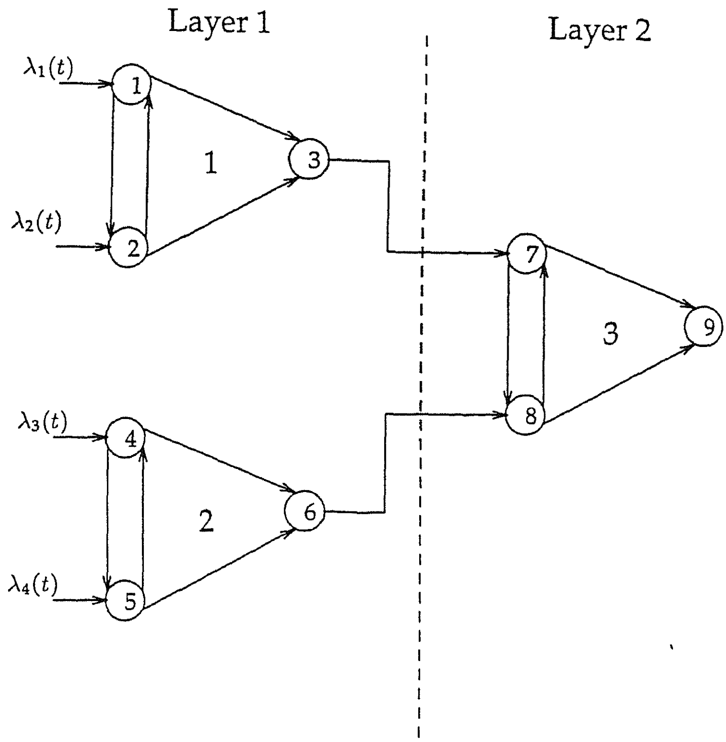

Figure 5 14· Network topology of three node units

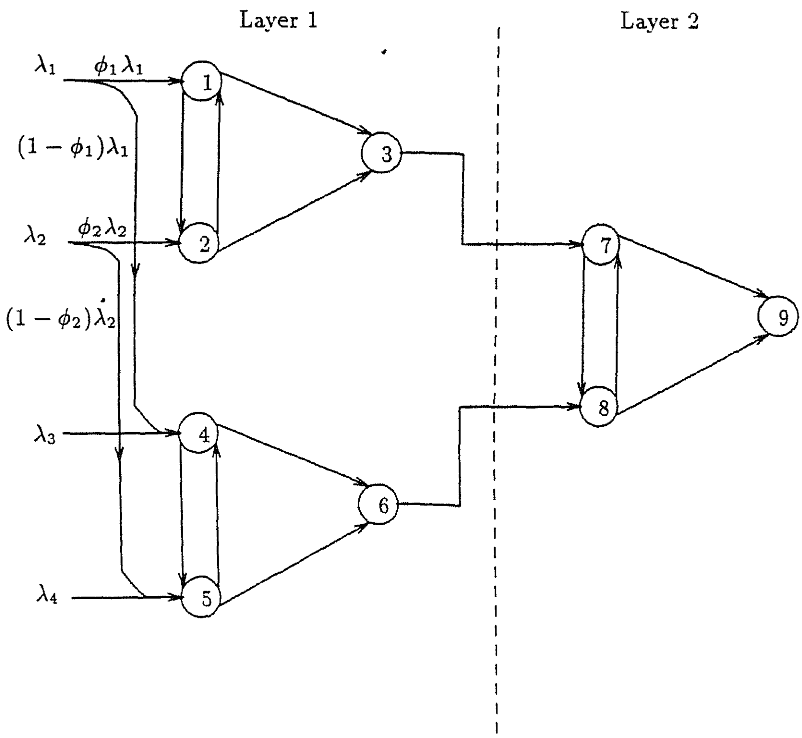

Figure 5 15 A scheme of traffic splitting for operating the network in the Linear

Table 5 5: Comparison of the performances of the *globally optimal* and *locally optimal* routing strategies for a load $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 10$

No	LINK PARAMETERS													
	a_{12}	a_{13}	a_{21}	a_{23}	a_{45}	a_{46}	a_{54}	a_{56}	a_{78}	a_{79}	a_{87}	a_{89}	J_1	J_2
1	0 2	0 5	0.2	0 5	0.2	0 6	0.2	0.6	0.1	0 5	0.1	0 5	1113.2	1113.2
2	0 9	0.2	0 1	0.8	0 2	0 6	0.2	0 6	0.1	0 5	0 1	0.5	972.16	1102.4
3	0 2	0.8	0 7	0.2	0 8	0.1	0.1	0 8	0.1	0 5	0 1	0 5	1103.2	1154 8
4	0.2	0 8	0 7	0 2	0 8	0 1	0 1	0.8	0 7	0 2	0.1	0 9	1107.6	1231 7
5	0 1	0.7	0.8	0 1	0 1	0 1	0 9	0.8	0.2	0 8	0 6	0.1	1161.8	1169 8

easily seen that the optimal routing strategy depends only on the link parameters a_{jk} of the network and is independent of the input loads $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_3(t)$ and $\lambda_4(t)$ As the number of layers in the network increases, the number of equations in the costate variables also increases For the topology shown in Figure 5 14 with two layers, in order to obtain the *globally optimal* routing strategy and the corresponding optimal buffer occupancies, the above set of 24 differential equations (Equations (5 8) to (5.31)) have to be solved Consider instead, the network performance under a scheme in which optimisation is done *local* to the individual units In Chapter 4, we have already looked at this problem and showed that such *locally optimal strategies* for each network unit can be specified by an inter-regime switching instant t_s which is easily obtainable from the link parameters of the unit. For a choice of $\lambda_1(t) = 10$, $\lambda_2(t) = 10$, $\lambda_3(t) = 10$ and $\lambda_4(t) = 10$ and link parameters a_{jk}, the Table 5 5 gives a quantitative difference in the performance of the network for the two schemes J_1 is the performance index under the *globally optimal routing scheme* and J_2 is the performance index under the *locally optimal routing scheme* The network operation is assumed to be for a duration of 10 units, i e $T = 10$ and the initial buffer occupancies are all assumed to be zero

The various cases in Table 5 5 are illustrative of the various modes in which the network operates under the two strategies Case 1 corresponds to a situation,

where both the *globally* optimal strategy and the *locally* optimal strategies are the same with a direct routing of the packets in all the three network units. In other words, the routing variables $\alpha_{13}(t)$, $\alpha_{23}(t)$ $\alpha_{46}(t)$, $\alpha_{56}(t)$, $\alpha_{79}(t)$, $\alpha_{89}(t)$ are unity in the entire duration of network operation The performance of the two strategies are same for this case.

Case 2 corresponds to a a situation in which there is a regime 2 to regime 4 transition in the network unit 1. In the network units 2 and 3, the incoming traffic is routed onto the direct paths for the entire duration $[0, T]$ under both the strategies. Under the *globally* optimal strategy, this regime transition (from 2 to 4) takes place at $t_s = 8\ 3$, while under the *locally* optimal strategy, this takes place at $t_s = 8\ 8$

Case 3 corresponds to a situation in which the network unit 1 has an inter regime switching from regime 3 to regime 4, while in the unit 2 there is an inter regime switching from regime 2 to regime 4. The routing in the network unit 3 is a direct routing for the entire duration $[0, T]$ Under the *globally* optimal scheme the inter regime switching in unit 1 takes place at $t_s = 7\ 7$ and in unit 2, at $t_s = 9\ 3$ Under the *locally* optimal scheme the inter regime switching in unit 1 takes place at $t_s = 8\ 2$ and the switching instant in the network unit 2 is at $t_s = 9\ 5$

Case 4 corresponds to a situation in which the network unit 1 has an inter regime switching from regime 3 to regime 4, unit 2 has a switching from regime 2 to regime 4, and unit 3 has a switching from regime 2 to regime 4 These instances are at $t_s = 7\ 3$, $t_s = 9\ 2$ and at $t_s = 8\ 4$ for the units 1, 2 and 3 respectively for the *globally* optimal strategy, while for the *locally* optimal strategies they are at $t_s = 8\ 2$, $t_s = 9\ 5$ and $t_s = 8\ 4$.

Case 5 corresponds to a situation in which all the three units have a inter regime switching from regime 3 to regime 4. These instances are at $t_s = 9\ 2$, $t_s = 8\ 3$ and at $t_s = 9\ 2$ for the units 1, 2 and 3 respectively for the *globally* optimal strategy, while for the *locally* optimal strategies they are at $t_s = 9\ 4$, $t_s = 8\ 8$ and $t_s = 9\ 2$ respectively

When a direct link of one of the basic units is of finite channel capacity, the network operation may go into saturation when the traffic on this link is high. Since the

maximum rate at which packets can be sent on this link is restricted to the channel capacity of the link and since for the rest of the links there is no upper bound on the rate at which packets can be drawn out, it is intuitively justifiable to redistribute the total input traffic in the network so that the network operates in the linear mode for the entire time span of network operation. Consider a situation in which link (1,3) of network unit 1 has a finite channel capacity A possible way to operate the network in the linear mode is to send the excess traffic (which causes the buffer occupancy in the link (1,3) increase beyond the value x_{13_s}) to the network unit 2 Figure (5.15) illustrates this scheme As mentioned above the routing strategy in the network can be either *globally* optimal or *locally* optimal for the constituent units and the network operation may be constrained to the linear mode with either of the routing strategies in operation Table 5 6 gives the performances under the *globally* optimal routing strategy and the *locally optimal* routing strategy (with the network operation being confined to the linear mode) for the above five cases of link parameters (given in Table 5 5) The input traffic is assumed to be $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = \lambda_4(t) = 10$ for the entire duration of network operation $[0, 10]$

5.4 Optimal routing in a four-node hub network

Consider a network topology shown in Figure 5.16 Nodes 1, 2, and 3 are source nodes and node 4 is the destination node At each source node, there is a direct path to the destination and alternate paths through the other source nodes. The above network can be viewed as being composed of 3 three-node network units of Chapter 4, wherein adjacent units share a common link. We investigate the nature of the optimal routing strategy for this network when all the links of the network are of infinite channel capacity.

Table 5 6 Performances of the *Globally optimal* and *locally optimal* strategies with the traffic splitting scheme

Case	C_{13}	Globally Optimal Scheme			Locally Optimal Scheme		
		ϕ_1	ϕ_2	J_1	ϕ_1	ϕ_2	J_2
1	5	$\phi_1(t) = 1, \forall t \in [0,1\ 4]$ $\phi_1(t) = 0\ 5,$ in $(1\ 4, 10]$	$\phi_2(t) = 1\ \forall t \in [0, 10]$	864 123	$\phi_1(t) = 1\ \forall t \in [0\ 1\ 4]$ $\phi_1(t) = 0\ 5$ in $(1\ 4, 10]$	$\phi_2(t) = 1 \forall t \in [0, 10]$	864 123
2	2	$\phi_1 = 1,$ in $[0.9\ 5]$ $\phi_1 = 0\ 2$ in $(9.5, 10]$	$\phi_2 = 1,$ in $[0\ 10]$	965 38	$\phi_1 = 1,$ in $[0, 9\ 8]$ $\phi_1 = 0\ 2$ in $(9\ 8, 10]$	$\phi_2 = 1$ in $[0, 10]$	979 38
3	8	$\phi_1 = 1,$in $[0, 1.13]$ $\phi_1 = e^{-0\ 7t} - 0.2$ in $(1\ 13, 2\ 2991]$ $\phi_1 = 0$ in $(2\ 2991, 7\ 6]$ $\phi_1 = 0\ 8(1 - e^{-0\ 7(t-7\ 6)})$ in $(7\ 6\ 10]$	$\phi_2 = 1,$ in $[0, 2\ 2991]$ $\phi_2 = 0\ 8$ in $(2.2991, 7\ 6]$ $= 1,$ in $(7\ 6, 10]$	825.35	$\phi_1 = 1,$ in $[0, 1.13]$ $\phi_1 = e^{-0\ 7t} - 0\ 2,$ in $(1\ 13, 2\ 2991]$ $\phi_2 == 0$ in $(2\ 2991, 8\ 2]$ $\phi_1 = 0\ 8(1 - e^{-0\ 7(t-8\ 2)})$ in $(8\ 2, 10]$	$\phi_2 = 1$ in $[0, 2.2991]$ $\phi_2 = 0.8$ in $(2.2991, 8\ 2]$ $\phi_2 = 1,$ in $(8\ 2, 10]$	879 86
4	8	$\phi_1 = 1$ in $[0.1\ 13]$ $\phi_1 = e^{-0\ 7t} - 0\ 2,$ in $(1\ 13, 2\ 2991]$ $\phi_1 = 0$ in $(2\ 2991, 7\ 3]$ $\phi_1 = 0\ 8(1 - e^{-0\ 7(t-7\ 3)})$ in $(7\ 3, 10]$	$\phi_2 = 1$ in $[0, 2\ 2991]$ $\phi_2 = 0\ 8$ in $(2\ 2991.7\ 3]$ $\phi_2 = 1$ in $(7\ 3, 10]$	1121 14	$\phi_1 = 1$ in $[0, 1\ 13]$ $\phi_1 = e^{-0\ 7t} - 0\ 2,$ in $(1.13, 2\ 2991]$ $\phi_1 = 0$ in $(2\ 2991, 8\ 2]$ $\phi_1 = 0\ 8(1 - e^{-0\ 7(t-8\ 2)})$ in $(8\ 2, 10]$	$\phi_2 = 1$ in $[0, 2, 2991]$ $\phi_2 = 0.8$ in $(2\ 2991, 8.2]$ $\phi_2 = 1$ in $(7\ 3, 10]$	1204 6
5	7	$\phi_1 = 1$ in $[0, 1\ 05]$ $\phi_1 = e^{-0\ 8t} - 0\ 3$ in $[1\ 05\ 1\ 505]$ $\phi_1 = 0$ in $(1\ 505, 9\ 2]$ $\phi_1 = 0\ 7(1 - e^{-0\ 8(t-9\ 2)}$ in $(9\ 2\ 10]$	$\phi_2 = 1$ in $[0, 1\ 505]$ $= 0\ 7$ in $(1\ 505, 9\ 2]$ $\phi_2 = 1$ in $(9\ 2, 10]$	1183 41	$\phi_1 = 1$ in $[0, 1\ 05]$ $\phi_1 = e^{-0\ 8t} - 0\ 3$ in $[1\ 05, 1\ 505]$ $\phi_1 = 0$ in $(1\ 505, 9\ 4]$ $\phi_1 = 0\ 7(1 - e^{-0\ 8(t-9\ 2)}$ in $(9\ 2, 10]$	$\phi_2 = 1$ in $[0, 1\ 50]$ $= 0\ 7$ in $(1\ 505, 9\ 2]$ $\phi_2 = 1$ in $(9\ 2, 10]$	1231 87

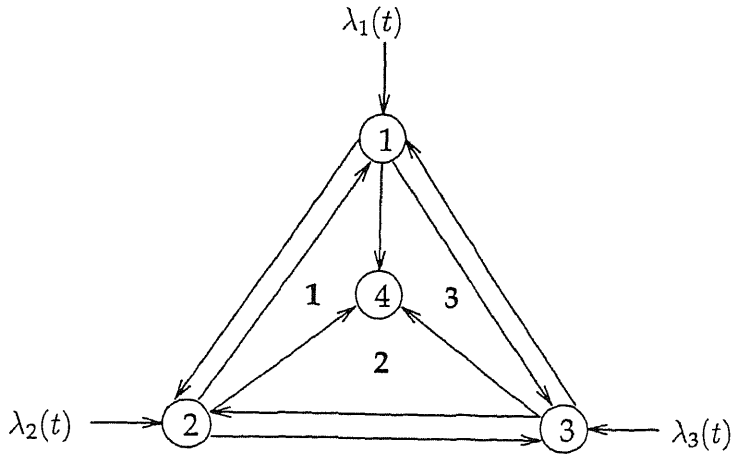

Figure 5.16. Four node hub network.

The dynamics of the network is given by the following set of equations

$$x_{12} = -a_{12}x_{12} + \alpha_{12}(t)(\lambda_1(t) + a_{21}x_{21} + a_{31}x_{31}) \tag{5.32}$$

$$x_{13} = -a_{13}x_{13} + \alpha_{13}(t)(\lambda_1(t) + a_{21}x_{21} + a_{31}x_{31}) \tag{5.33}$$

$$x_{14} = -a_{14}x_{14} + \alpha_{14}(t)(\lambda_1(t) + a_{21}x_{21} + a_{31}x_{31}) \tag{5.34}$$

$$x_{21} = -a_{21}x_{21} + \alpha_{21}(t)(\lambda_2(t) + a_{12}x_{12} + a_{32}x_{32}) \tag{5.35}$$

$$x_{23} = -a_{23}x_{23} + \alpha_{23}(t)(\lambda_2(t) + a_{12}x_{12} + a_{32}x_{32}) \tag{5.36}$$

$$x_{24} = -a_{24}x_{24} + \alpha_{24}(t)(\lambda_2(t) + a_{12}x_{12} + a_{32}x_{32}) \tag{5.37}$$

$$x_{31} = -a_{31}x_{31} + \alpha_{31}(t)(\lambda_3(t) + a_{13}x_{13} + a_{23}x_{23}) \tag{5.38}$$

$$x_{32} = -a_{32}x_{32} + \alpha_{32}(t)(\lambda_3(t) + a_{13}x_{13} + a_{23}x_{23}) \tag{5.39}$$

$$x_{34} = -a_{34}x_{34} + \alpha_{34}(t)(\lambda_3(t) + a_{13}x_{13} + a_{23}x_{23}) \tag{5.40}$$

Consider a routing strategy which minimises the total buffer occupancy time J given below

$$J = \int_0^T (x_{12} + x_{13} + x_{14} + x_{21} + x_{23} + x_{24} + x_{31} + x_{32} + x_{34})dt$$

The Hamiltonian for this system is given as

$$H = x_{12} + x_{13} + x_{14} + x_{21} + x_{23} + x_{24} + x_{31} + x_{32} + x_{34}$$

$$+p_{12}x_{12} + p_{13}\dot{x}_{13} + p_{14}\dot{x}_{14} + p_{21}x_{21} + p_{23}x_{23} + p_{24}x_{24}$$
$$+p_{31}x_{31} + p_{32}\dot{x}_{32} + p_{34}x_{34}$$

Substituting the dynamics given by Equations (5.32)-(5.40) we get

$$\begin{aligned}
H &= x_{12} + x_{13} + x_{14} + x_{21} + x_{23} + x_{24} + x_{31} + x_{32} + x_{34} \\
&\quad +p_{12}(-a_{12}x_{12} + \alpha_{12}(\lambda_1 + a_{21}x_{21} + a_{31}x_{31})) \\
&\quad +p_{13}(-a_{13}x_{13} + \alpha_{13}(\lambda_1 + a_{21}x_{21} + a_{31}x_{31})) \\
&\quad +p_{14}(-a_{14}x_{14} + \alpha_{14}(\lambda_1 + a_{21}x_{21} + a_{31}x_{31})) \\
&\quad +p_{21}(-a_{21}x_{21} + \alpha_{21}(\lambda_2 + a_{12}x_{12} + a_{13}x_{13})) \\
&\quad +p_{23}(-a_{23}x_{23} + \alpha_{23}(\lambda_2 + a_{12}x_{12} + a_{32}x_{32})) \\
&\quad +p_{24}(-a_{24}x_{24} + \alpha_{24}(\lambda_2 + a_{12}x_{12} + a_{32}x_{32})) \\
&\quad +p_{31}(-a_{31}x_{31} + \alpha_{31}(\lambda_3 + a_{13}x_{13} + a_{23}x_{13})) \\
&\quad +p_{32}(-a_{32}x_{32} + \alpha_{32}(\lambda_3 + a_{13}x_{13} + a_{23}x_{13})) \\
&\quad +p_{34}(-a_{34}x_{34} + \alpha_{34}(\lambda_3 + a_{13}x_{13} + a_{23}x_{13}))
\end{aligned}$$

Minimising the Hamiltonian function w r.t the routing variables is equivalent to minimising the expression

$$\begin{aligned}
&(p_{12}\alpha_{12} + p_{13}\alpha_{13} + p_{14}\alpha_{14})(\lambda_1 + a_{21}x_{21} + a_{31}x_{31}) \\
&+(p_{21}\alpha_{21} + p_{23}\alpha_{23} + p_{24}\alpha_{24})(\lambda_2 + a_{12}x_{12} + a_{32}x_{32}) \\
&+(p_{31}\alpha_{31} + p_{32}\alpha_{32} + p_{34}\alpha_{34})(\lambda_3 + a_{13}x_{13} + a_{23}x_{23})
\end{aligned}$$

Since the terms $(\lambda_1 + a_{21}x_{21} + a_{31}x_{31})$, $(\lambda_2 + a_{12}x_{12} + a_{32}x_{32})$ and $(\lambda_3 + a_{13}x_{13} + a_{23}x_{23})$ are non-negative the optimal routing variables are specified as follows.

- The optimal routing variables $\alpha_{12}(t)$, $\alpha_{13}(t)$ and $\alpha_{14}(t)$ where $0 \leq \alpha_{12}(t) \leq 1$, $0 \leq \alpha_{13}(t) \leq 1$, $0 \leq \alpha_{14}(t) \leq 1$ and $\alpha_{12}(t) + \alpha_{13}(t) + \alpha_{14}(t) = 1$ are such that they minimise the term $(p_{12}\alpha_{12} + p_{13}\alpha_{13} + p_{14}\alpha_{14})$

- The optimal routing variables $\alpha_{21}(t)$, $\alpha_{23}(t)$ and $\alpha_{24}(t)$ are such that they minimise the term $(p_{21}\alpha_{21} + p_{23}\alpha_{23} + p_{24}\alpha_{24})$

- The optimal routing variables $\alpha_{31}(t)$, $\alpha_{32}(t)$ and $\alpha_{34}(t)$ are such that they minimise the term $(p_{31}\alpha_{31} + p_{32}\alpha_{32} + p_{34}\alpha_{34})$

Let $\hat{p}_1(t)$ be the minimum of $p_{12}(t)$, $p_{13}(t)$ and $p_{14}(t)$. Then for $j \in \{2,3,4\}$

$$\alpha_{1j}(t) = \begin{cases} 1 & \text{if } p_{1j}(t) = \hat{p}_1(t) \\ 0 & \text{otherwise} \end{cases}$$

Similarly, let $\hat{p}_2(t)$ be the minimum of $p_{21}(t)$, $p_{23}(t)$ and $p_{24}(t)$ Then for $j \in \{1,3,4\}$

$$\alpha_{2j}(t) = \begin{cases} 1 & \text{if } p_{2j}(t) = \hat{p}_2(t) \\ 0 & \text{otherwise} \end{cases}$$

And, let $\hat{p}_3(t)$ be the minimum of $p_{31}(t)$, $p_{32}(t)$ and $p_{34}(t)$ Then for $j \in \{1,2,4\}$

$$\alpha_{3j}(t) = \begin{cases} 1 & \text{if } p_{3j}(t) = \hat{p}_3(t) \\ 0 & \text{otherwise} \end{cases}$$

The costate variables are $p_{jk}(t)$'s are given by the following differential equations:

$$p_{12} = -1 + a_{12}(p_{12} - \alpha_{21}p_{21} - \alpha_{23}p_{23} - \alpha_{24}p_{24}) \tag{5.41}$$

$$p_{13} = -1 + a_{13}(p_{13} - \alpha_{31}p_{31} - \alpha_{32}p_{32} - \alpha_{34}p_{34}) \tag{5.42}$$

$$p_{14} = -1 + a_{14}p_{14} \tag{5.43}$$

$$p_{21} = -1 + a_{21}(p_{21} - \alpha_{12}p_{12} - \alpha_{13}p_{13} - \alpha_{14}p_{14}) \tag{5.44}$$

$$p_{23} = -1 + a_{23}(p_{23} - \alpha_{31}p_{31} - \alpha_{32}p_{32} - \alpha_{34}p_{34}) \tag{5.45}$$

$$p_{24} = -1 + a_{24}p_{24} \tag{5.46}$$

$$p_{31} = -1 + a_{31}(p_{31} - \alpha_{12}p_{12} - \alpha_{13}p_{13} - \alpha_{14}p_{14}) \tag{5.47}$$

$$p_{32} = -1 + a_{32}(p_{32} - \alpha_{21}p_{21} - \alpha_{23}p_{23} - \alpha_{24}p_{24}) \tag{5.48}$$

$$p_{34} = -1 + a_{34}p_{34} \tag{5.49}$$

along with the transversality conditions $p_{jk}(T) = 0$ The solutions for $p_{14}(t)$, $p_{24}(t)$ and $p_{34}(t)$ are given as

$$p_{14}(t) = \frac{1 - e^{a_{14}(t-T)}}{a_{14}}$$

$$p_{24}(t) = \frac{1-e^{a_{24}(t-T)}}{a_{24}}$$
$$p_{34}(t) = \frac{1-e^{a_{34}(t-T)}}{a_{34}}$$

Remark 5.4.1 : It can be argued from the above set of differential equations in the costate variables that there do not exist intervals during which $p_{12}(t) \equiv p_{14}(t)$ and/or $p_{13}(t) \equiv p_{14}(t)$ Similarly intervals during which $p_{21}(t) \equiv p_{24}(t)$ and/or $p_{23}(t) \equiv p_{24}(t)$ and/or $p_{31}(t) \equiv p_{34}(t)$ and/or $p_{32}(t) \equiv p_{34}(t)$ also do not exist. If during an interval I, $p_{12}(t) \equiv p_{13}(t) \equiv \hat{p}_1(t)$ (a case corresponding to singular solutions) then it can be argued that during I, $\alpha_{14}(t) \equiv 0$ and any choice of $\alpha_{12}(t)$ and $\alpha_{13}(t)$ which satisfies the normalizing conditions is an optimal strategy This observation applies to other cases also. Since a choice of unity for one of the (associated) routing variables under these situations is also an optimal choice, in the discussion that follows we restrict our attention to the class of *bang-bang* strategies.

Since the optimal routing strategy at each source node is *bang-bang* with one of the associated routing variables always taking the value equal to unity, we have the following 27 possible regimes of operation which are given in Table 5 7. The optimal routing strategy for this four node network has certain interesting properties which are given below as theorems.

Theorem 5.4.1 *The network operation always ends in regime 27 wherein the incoming packets at the source nodes are routed on the direct links (1,4), (2,4) and (3,4)*

Proof:

We prove that the rest of the regimes are ruled out as terminal regimes as follows.

Regimes 1,2 and 3 : In all these three regimes, the routing variables $\alpha_{12}(t)$ and $\alpha_{21}(t)$ are equal to unity The dynamics of $p_{12}(t)$ and $p_{21}(t)$ are given as

$$p_{12}(t) = -1 + a_{12}(p_{12} - p_{21})$$
$$p_{21}(t) = -1 + a_{21}(p_{21} - p_{12})$$

Table 5 7. Possible regimes of network operation

Regime no.	α_{12}	α_{13}	α_{14}	α_{21}	α_{23}	α_{24}	α_{31}	α_{32}	α_{34}
Regime 1	1	0	0	1	0	0	1	0	0
Regime 2	1	0	0	1	0	0	0	1	0
Regime 3	1	0	0	1	0	0	0	0	1
Regime 4	1	0	0	0	1	0	1	0	0
Regime 5	1	0	0	0	1	0	0	1	0
Regime 6	1	0	0	0	1	0	0	0	1
Regime 7	1	0	0	0	0	1	1	0	0
Regime 8	1	0	0	0	0	1	0	1	0
Regime 9	1	0	0	0	0	1	0	0	1
Regime 10	0	1	0	1	0	0	1	0	0
Regime 11	0	1	0	1	0	0	0	1	0
Regime 12	0	1	0	1	0	0	0	0	1
Regime 13	0	1	0	0	1	0	1	0	0
Regime 14	0	1	0	0	1	0	0	1	0
Regime 15	0	1	0	0	1	0	0	0	1
Regime 16	0	1	0	0	0	1	1	0	0
Regime 17	0	1	0	0	0	1	0	1	0
Regime 18	0	1	0	0	0	1	0	0	1
Regime 19	0	0	1	1	0	0	1	0	0
Regime 20	0	0	1	1	0	0	0	1	0
Regime 21	0	0	1	1	0	0	0	0	1
Regime 22	0	0	1	0	1	0	1	0	0
Regime 23	0	0	1	0	1	0	0	1	0
Regime 24	0	0	1	0	1	0	0	0	1
Regime 25	0	0	1	0	0	1	1	0	0
Regime 26	0	0	1	0	0	1	0	1	0
Regime 27	0	0	1	0	0	1	0	0	1

with the transversality conditions $p_{12}(T) = p_{21}(T) = 0$

The solution to the above equations are

$$
\begin{aligned}
p_{12}(t) &= (T-t) \\
p_{21}(t) &= (T-t)
\end{aligned}
$$

Since $p_{12}(t) = T - t > p_{14}(t) = \frac{1-e^{a_{14}(t-T)}}{a_{14}}$, the routing variable $\alpha_{12}(t)$ can not be unity Similarly $p_{21}(t) = T - t > p_{24}(t) = \frac{1-e^{a_{24}(t-T)}}{a_{24}}$ and therefore the routing variable $\alpha_{21}(t)$ can not be unity.

Regime 5 :

In regime 5, the routing variables $\alpha_{12}(t) = 1$, $\alpha_{23}(t) = 1$, and $\alpha_{32}(t) = 1$ The dynamics of $p_{23}(t)$ and $p_{32}(t)$ are given as

$$
\begin{aligned}
p_{23} &= -1 + a_{23}(p_{23} - p_{32}) \\
p_{32} &= -1 + a_{32}(p_{32} - p_{23})
\end{aligned}
$$

with the transversality condition $p_{23}(T) = p_{32}(T) = 0$

Thus $p_{23}(t) = p_{32}(t) = T - t$. Since $p_{23}(t) = (T - t) > p_{24}(t) = \frac{1-e^{a_{24}(t-T)}}{a_{24}}$, $\alpha_{23}(t)$ can not be unity. Similarly $\alpha_{32}(t)$ can also not be unity Thus regime 5 is ruled out as a terminal regime.

Regime 6 :

$\alpha_{12} = 1$, $\alpha_{23} = 1$ and $\alpha_{31} = 1$ The dynamics of $p_{12}(t)$ and $p_{23}(t)$ are given as

$$
\begin{aligned}
p_{12} &= -1 + a_{12}(p_{12} - p_{23}) \\
p_{23} &= -1 + a_{23}(p_{23} - p_{34})
\end{aligned}
$$

with the transversality condition $p_{12}(T) = p_{23}(T) = 0$

The solution for $p_{23}(t)$ is

$$
p_{23}(t) = \frac{1 - e^{a_{23}(t-T)}}{a_{23}} + \frac{1 - e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{23}(t-T)} - e^{a_{34}(t-T)}}{a_{23} - a_{34}}
$$

As proved in Theorem 4 3 3, the function $p_{23}(t)$ as given above is greater than $p_{24}(t)$ in some interval ending with T Therefore the routing variable $\alpha_{23}(t)$ can not be unity and regime 6 as a terminal regime is ruled out

Regimes 7, 8 and 9 :

$\alpha_{12} = 1$, $\alpha_{24} = 1$.

The dynamics of $p_{12}(t)$ is given as

$$p_{12} = -1 + a_{12}(p_{12} - p_{24})$$

with the transversality condition $p_{12}(T) = 0$.

The solution to the above is given as

$$p_{12}(t) = \frac{1 - e^{a_{12}(t-T)}}{a_{12}} + \frac{1 - e^{a_{24}(t-T)}}{a_{24}} + \frac{e^{a_{12}(t-T)} - e^{a_{24}(t-T)}}{a_{12} - a_{24}}$$

Since $p_{12}(t)$ as given by the above expression can not be less than $p_{14}(t)$ in an interval ending with T, the routing variable $\alpha_{12}(t)$ can not be unity Hence regimes 7, 8 and 9 are ruled out as terminal regimes.

Regime 10 :

$\alpha_{13}(t) = 1$, $\alpha_{31}(t) = 1$.

The dynamics of $p_{13}(t)$ and $p_{31}(t)$ are given as

$$p_{13} = -1 + a_{13}(p_{13} - p_{31})$$
$$p_{31} = -1 + a_{31}(p_{31} - p_{13})$$

with the transversality condition $p_{13}(T) = p_{31}(T) = 0$.

The solution to the above is given as

$$p_{13}(t) = (T - t)$$
$$p_{31}(t) = (T - t)$$

Since $p_{13}(t) = T - t > p_{14}(t)$, the routing variable $\alpha_{13}(t)$ can not be unity. Similarly the routing variable $\alpha_{31}(t)$ can not be unity. Therefore regime 10 as a terminal regime is ruled out.

Regime 11 :

The routing variables $\alpha_{13}(t)$, $\alpha_{21}(t)$ and $\alpha_{32}(t)$ are unity in this case. The dynamics of $p_{13}(t)$, $p_{21}(t)$ and $p_{32}(t)$ are given as

$$p_{13} = -1 + a_{13}(p_{13} - p_{32})$$

$$\begin{aligned} p_{21} &= -1 + a_{21}(p_{21} - p_{13}) \\ p_{32} &= -1 + a_{32}(p_{32} - p_{21}) \end{aligned}$$

with the transversality conditions $p_{13}(T) = p_{21}(T) = p_{32}(T) = 0$
The solutions to the above are $p_{13}(t) = p_{21}(t) = p_{32}(t) = T - t$. Since $p_{13}(t)$ is greater than $p_{14}(t)$, the routing variable $\alpha_{13}(t)$ can not assume a value equal to unity This holds true for the routing variables $\alpha_{21}(t)$ and $\alpha_{32}(t)$ too Thus regime 11 as a terminal regime is ruled out.

Regime 12 :

The routing variables $\alpha_{13}(t)$, $\alpha_{21}(t)$ and $\alpha_{34}(t)$ are unity. In regime 12, the costate variable $p_{13}(t)$ satisfies the following differential equation

$$p_{13}(t) = -1 + a_{13}(p_{13} - p_{34})$$

with the transversality condition $p_{13}(T) = 0$
The solution to the above equation is

$$p_{13}(t) = \frac{1 - e^{a_{13}(t-T)}}{a_{13}} + \frac{1 - e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{13}(t-T)} - e^{a_{34}(t-T)}}{a_{13} - a_{34}}$$

Since the routing variable $p_{13}(t)$ is greater than $p_{14}(t)$ in an interval ending with T, the routing variable $\alpha_{13}(t)$ can not assume the value unity. Hence regime 12 as a terminal regime is ruled out.

Regimes 13 and 16 :

$\alpha_{13}(t)$ and $\alpha_{31}(t)$ are unity in these regimes The costate variables $p_{13}(t)$ and $p_{31}(t)$ satisfy the following differential equations

$$\begin{aligned} p_{13} &= -1 + a_{13}(p_{13} - p_{31}) \\ p_{31} &= -1 + a_{31}(p_{31} - p_{13}) \end{aligned}$$

with the terminal conditions $p_{13}(T) = p_{31}(T) = 0$
$p_{13}(t) = p_{31}(t) = T - t$ As argued earlier, $p_{13}(t)$ ($p_{31}(t)$) is greater than $p_{31}(t)$ ($p_{34}(t)$).

Hence the choice of the value for the routing variables $\alpha_{13}(t)$ and $\alpha_{31}(t)$ as unity in an interval ending with T results in the violation of the optimality condition. Hence regime 13 and 16 are ruled out as terminal regimes

Regime 14 :

In regime 14, the routing variables α_{23} and $\alpha_{32}(t)$ are unity. By an analogous argument as given above in the routing variables $\alpha_{23}(t)$ and $\alpha_{32}(t)$ and the costate variables $p_{23}(t)$ and $p_{32}(t)$ results in the fact that the choice of the value of unity for $\alpha_{23}(t)$ and $\alpha_{32}(t)$ in an interval ending with T violates the optimality condition. Thus regime 14 can not be a terminal regime.

Regime 15 and 18 :

In regime 15 and 18, the routing variables $\alpha_{13}(t)$ and $\alpha_{31}(t)$ are unity. The dynamics of $p_{13}(t)$ is given as

$$p_{13} = -1 + a_{13}(p_{13} - p_{34})$$

with the transversality condition $p_{13}(T) = 0$

The solution to the above equation is given as

$$p_{13}(t) = \frac{1 - e^{a_{13}(t-T)}}{a_{13}} + \frac{1 - e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{13}(t-T)} - e^{a_{34}(t-T)}}{a_{13} - a_{34}}$$

Since $p_{13}(t)$ as given by the above expression is greater than $p_{14}(t)$ in an interval ending with T, the choice of the value for the routing variable $\alpha_{13}(t)$ as unity violates the optimality condition. Hence the regimes 15 and 18 are ruled out as terminal regimes

Regime 17 :

The routing variables $\alpha_{13}(t)$, $\alpha_{24}(t)$ and $\alpha_{32}(t)$ are unity The dynamics of $p_{32}(t)$ in regime 17 ending with T is given as

$$p_{32}(t) = \frac{1 - e^{a_{32}(t-T)}}{a_{32}} + \frac{1 - e^{a_{24}(t-T)}}{a_{24}} + \frac{e^{a_{32}(t-T)} - e^{a_{24}(t-T)}}{a_{32} - a_{24}}$$

The function $p_{32}(t)$ as given above is greater than $p_{34}(t)$ in an interval ending with T and therefore the routing variable $\alpha_{32}(t)$ can not be unity. Hence regime 17 is ruled out as a terminal regime.

Regimes 19, 20 and 21 :

In regimes 19, 20 and 21, the routing variables $\alpha_{21}(t)$ and $\alpha_{14}(t)$ are unity. The dynamics of $p_{21}(t)$ is given as

$$p_{21}(t) = \frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{a_{21} - a_{14}}$$

As argued earlier, the function $p_{21}(t)$ as given by the above expression is greater than $p_{24}(t)$ in an interval ending with T, the routing variable $\alpha_{21}(t)$ can not be unity. Hence the network operation can not end in regimes 19, 20 and 21 under the optimal routing strategy.

Regime 22 :

In regime 22, $\alpha_{14}(t)$ and $\alpha_{31}(t)$ are unity The dynamics of $p_{31}(t)$ is given as

$$p_{31}(t) = \frac{1 - e^{a_{31}(t-T)}}{a_{31}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{31}(t-T)} - e^{a_{14}(t-T)}}{a_{31} - a_{14}}$$

Since the function $p_{31}(t)$ as given by the above expression is greater than $p_{34}(t)$ in an interval ending with T, the routing variable $\alpha_{31}(t)$ can not be unity. Hence regime 22 is ruled out as a terminal regime.

Regime 23 :

The routing variables $\alpha_{23}(t)$ and $\alpha_{32}(t)$ are unity in regime 23. It is easy to verify that the costate variables $p_{23}(t)$ and $p_{32}(t)$ are given as

$$p_{23}(t) = p_{32}(t) = T - t$$

Since $p_{23}(t)$ ($p_{32}(t)$) is greater than $p_{24}(t)$ ($p_{34}(t)$) the choice of the value for routing variable $\alpha_{23}(t)$ ($\alpha_{32}(t)$) as unity violates the optimality condition.

Regime 24 :

$\alpha_{23}(t)$ and $\alpha_{34}(t)$ are unity in regime 24

The costate variable $p_{23}(t)$ is given as

$$p_{23}(t) = \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{1-e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{23}(t-T)} - e^{a_{34}(t-T)}}{a_{23} - a_{34}}$$

Since $p_{23}(t)$ is greater than $p_{24}(t)$ in an interval ending with T, the routing variable $\alpha_{23}(t)$ can not be unity. Thus regime 24 is ruled out as a terminal regime.

Regime 25 :

In regime 25, α_{14} and α_{31} are unity The costate variable $p_{31}(t)$ is given as

$$p_{31}(t) = \frac{1-e^{a_{31}(t-T)}}{a_{31}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{31}(t-T)} - e^{a_{14}(t-T)}}{a_{31} - a_{14}}$$

Since $p_{31}(t)$ is greater than $p_{34}(t)$ in an interval ending with T, the routing variable $\alpha_{31}(t)$ can not be unity Thus regime 25 is ruled out as a terminal regime

Regime 26 :

In regime 26, α_{32} and α_{24} are unity The costate variable $p_{32}(t)$ is given as

$$p_{32}(t) = \frac{1-e^{a_{32}(t-T)}}{a_{32}} + \frac{1-e^{a_{24}(t-T)}}{a_{24}} + \frac{e^{a_{32}(t-T)} - e^{a_{24}(t-T)}}{a_{32} - a_{24}}$$

Since $p_{32}(t)$ is greater than $p_{34}(t)$ in an interval ending with T, the routing variable $\alpha_{32}(t)$ can not be unity Thus regime 26 is ruled out as a terminal regime.

□

Thus the network operation always ends in regime 27 under the optimal routing strategy. In a regime 27 ending with T, the costate variables are given as the following.

$$p_{12}(t) = \frac{1-e^{a_{12}(t-T)}}{a_{12}} + \frac{1-e^{a_{24}(t-T)}}{a_{24}} + \frac{e^{a_{12}(t-T)} - e^{a_{24}(t-T)}}{a_{12} - a_{24}} \tag{5 50}$$

$$p_{13}(t) = \frac{1-e^{a_{13}(t-T)}}{a_{13}} + \frac{1-e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{13}(t-T)} - e^{a_{34}(t-T)}}{a_{13} - a_{34}} \tag{5 51}$$

$$p_{14}(t) = \frac{1-e^{a_{14}(t-T)}}{a_{14}} \tag{5.52}$$

$$p_{21}(t) = \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{a_{21}-a_{14}} \quad (5\ 53)$$

$$p_{23}(t) = \frac{1-e^{a_{23}(t-T)}}{a_{23}} + \frac{1-e^{a_{34}(t-T)}}{a_{34}} + \frac{e^{a_{23}(t-T)} - e^{a_{34}(t-T)}}{a_{23}-a_{34}} \quad (5.54)$$

$$p_{24}(t) = \frac{1-e^{a_{24}(t-T)}}{a_{24}} \quad (5\ 55)$$

$$p_{31}(t) = \frac{1-e^{a_{31}(t-T)}}{a_{31}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{31}(t-T)} - e^{a_{14}(t-T)}}{a_{31}-a_{14}} \quad (5\ 56)$$

$$p_{32}(t) = \frac{1-e^{a_{32}(t-T)}}{a_{32}} + \frac{1-e^{a_{24}(t-T)}}{a_{24}} + \frac{e^{a_{32}(t-T)} - e^{a_{24}(t-T)}}{a_{32}-a_{24}} \quad (5\ 57)$$

$$p_{34}(t) = \frac{1-e^{a_{34}(t-T)}}{a_{34}} \quad (5.58)$$

Consider the situation where in all the direct links in the network have the same value for the link parameter. In practical terms, this amounts to the fact that the rate at which the links operates are the same for all the three direct links An interesting property of the optimal routing strategy is that under this situation, all the traffic arriving at the source nodes are routed onto the direct links and the alternate paths are never used for the entire duration of network operation

Theorem 5.4.2 *If all the direct links of the network have the same value for the link parameter, then the network operates in regime 27 for the entire duration* $[0, T]$

Proof :

If $a_{14} = a_{24} = a_{34}$, then $p_{14}(t) = p_{24}(t) = p_{34}(t)\ \forall t \in [0, T]$ Corresponding to this situation the costate variables as given by the equations (5 50)-(5 58) satisfy the following relationships

$$\begin{aligned}
p_{14}(t) &< p_{12}(t), \quad \forall t \in [0,T] \\
p_{14}(t) &< p_{13}(t), \quad \forall t \in [0,T]. \\
p_{24}(t) &< p_{21}(t), \quad \forall t \in [0,T] \\
p_{24}(t) &< p_{23}(t), \quad \forall t \in [0,T] \\
p_{34}(t) &< p_{31}(t), \quad \forall t \in [0,T] \\
p_{34}(t) &< p_{32}(t), \quad \forall t \in [0,T]
\end{aligned}$$

Hence the network operates in regime 27 for the entire duration $[0, T]$.

□

Consider the case wherein the direct links of the network are of different parameters. Let us assume that the link (1,4) is the *fastest* of the three In other words $a_{14} \geq max\{a_{24}, a_{34}\}$ Then it can be easily argued that the terminal regime 27 can not be preceded by a regime in which the link (1,4) is not used for routing the incoming packets at node 1.

Theorem 5.4.3 *Let $a_{14} \geq max\{a_{24}, a_{34}\}$ Then the terminal regime 27 can not be preceded by regimes 1 to 18*

Proof :
Since $a_{14} \geq max\{a_{24}, a_{34}\}$, it can be easily verified that the following relationships hold true for all $t \in [0, T]$.

$$\begin{aligned} p_{14}(t) &\leq \frac{1 - e^{a_{24}(t-T)}}{a_{24}} \\ p_{14}(t) &\leq \frac{1 - e^{a_{34}(t-T)}}{a_{34}} \end{aligned}$$

Since $p_{12}(t)$ and $p_{13}(t)$ as given by the equations (5 50) and (5.51) in a terminal regime 27 are greater than $p_{24}(t)$ and $p_{34}(t)$ respectively for all $t \in [0, T]$ (as argued in the proof of Theorem 4 3 2), they are both greater than $p_{14}(t), \forall t \in [0, T)$ For the terminal regime 27 to be preceded by any regime in which $\alpha_{14}(t)$ is equal to zero, either $p_{14}(t)$ has to be greater than $p_{12}(t)$ or $p_{14}(t)$ greater than $p_{13}(t)$ or possibly both. Since this can not be the case under the condition stated in the theorem, regime 27 can not be preceded by regimes 1 to 18, wherein $\alpha_{14}(t)$ equals zero Thus the only regimes that could possibly precede the regime 27 are 19, 20, 21, 22, 23, 24, 25 and 26.

Remark 5.4.2 : Reasoning exactly along the above lines, it can be concluded that when link (2,4) is the *fastest* direct link, then the terminal regime 27 can not be preceded by regimes 1, 2, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 19, 20, 21, 22, 23 and 24 Similarly if link (3,4) is the *fastest* direct link, then the terminal regime 27 can

not be preceded by regimes 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25 and 26.

We now investigate in the next subsection, the possible inter regime transitions and the procedure to specify these regime transition instants for the case wherein link (1,4) is the *fastest* direct link. From the topological symmetry of the network, it can be seen that the arguments used in the next subsection can be easily extended to the cases wherein either link (2,4) or (3,4) is the *fastest* direct link. This being the case, the arguments for the latter cases are not repeated

5.4.1 Links (2,4) and (3,4) of the same link parameter

Consider the case in which the direct links (2,4) and (3,4) have the same link parameter and are both *slower* than (1,4). In other words, $a_{24} = a_{34} < a_{14}$
In the terminal regime 27 the dynamics of $p_{23}(t)$ and $p_{32}(t)$ as given by Equations (5.54) and (5.57) can be written as the following :

$$
\begin{aligned}
p_{23}(t) &= p_{24}(t) + f_{23}(t) \\
p_{32}(t) &= p_{34}(t) + f_{32}(t)
\end{aligned}
$$

where the functions $f_{23}(t)$ and $f_{32}(t)$ are non-negative Therefore regime 27 can not be preceded by any regime in which $\alpha_{23}(t)$ and/or $\alpha_{32}(t)$ are unity Thus out of the 8 regimes 19 to 26 listed above, the ones that could possibly precede the regime 27 are 19, 21 and 25 only as shown in the three cases A, B and C in Figure 5.17 Let us examine the conditions on the link parameters of the network under which these transitions take place, and also the implications of these conditions

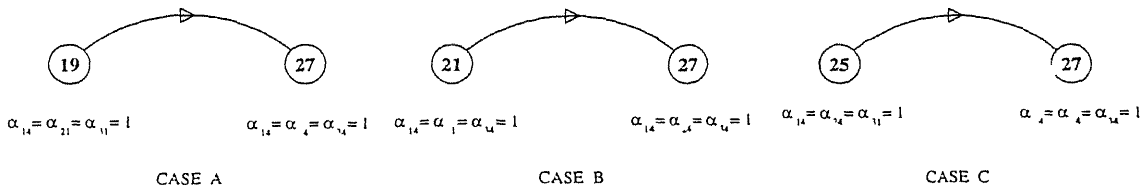

Figure 5.17 Possible regimes preceding regime 27 when $a_{24} = a_{34} < a_{14}$

CASE A :

Corresponding to the condition $a_{24} = a_{34}$, the costate variables $p_{24}(t)$, $p_{34}(t)$, $p_{21}(t)$ and $p_{31}(t)$ in terminal regime 27 are given as the following :

$$p_{24}(t) = p_{34}(t) = \frac{1 - e^{a_{24}(t-T)}}{a_{24}} \tag{5 59}$$

$$p_{21}(t) = \frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{(a_{21} - a_{14})} \tag{5.60}$$

$$p_{31}(t) = \frac{1 - e^{a_{31}(t-T)}}{a_{31}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{31}(t-T)} - e^{a_{14}(t-T)}}{(a_{31} - a_{14})} \tag{5 61}$$

For the terminal regime 27 to be preceded by regime 19 (in which $\alpha_{21} = \alpha_{31} = 1$) at some instant t_{s_1}, it is necessary and sufficient that

$$p_{21}(t_{s_1}) = p_{24}(t_{s_1})$$
$$p_{31}(t_{s_1}) = p_{34}(t_{s_1})$$

Since $p_{24}(t) = p_{34}(t)$, the above two conditions can be satisfied only if $p_{21}(t_{s_1}) = p_{31}(t_{s_1})$ **For this to be the case it is necessary and sufficient that** $a_{21} = a_{31}$ For the condition $p_{21}(t_{s_1}) = p_{24}(t_{s_1})$ to be satisfied for some $0 < t_{s_1} < T$, it is necessary and sufficient that

$$p_{21}(0) = \frac{1 - e^{-a_{21}T}}{a_{21}} + \frac{1 - e^{-a_{14}T}}{a_{14}} + \frac{(e^{-a_{21}T} - e^{-a_{14}T})}{(a_{21} - a_{14})} < p_{24}(0) = \frac{1 - e^{-a_{24}T}}{a_{24}}$$

The switching instant t_{s_1} is obtained by solving the equation

$$\frac{1 - e^{a_{21}(t_{s_1}-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t_{s_1}-T)}}{a_{14}} + \frac{(e^{a_{21}(t_{s_1}-T)} - e^{a_{14}(t_{s_1}-T)})}{(a_{21} - a_{14})} = \frac{1 - e^{a_{24}(t_{s_1}-T)}}{a_{24}}$$

It can be argued along the following lines that the regime 19 which precedes regime 27 has to be the starting regime

Note that the functions $p_{21}(t)$ and $p_{24}(t)$ can have at the most one point of intersection in the interval $(-\infty, T)$ (as proved in the Lemma 4 3 1) Same is the case with the functions $p_{31}(t)$ and $p_{34}(t)$ also. Therefore $p_{21}(t) < p_{24}(t)$ and $p_{31}(t) < p_{34}(t)$, $\forall t \in [0, t_{s_1}]$ The relationships $p_{21}(t) < p_{23}(t)$ and $p_{31}(t) < p_{32}(t)$ also hold true for all $t \leq t_{s_1}$ as can be justified on the following grounds:

In regime 19,

$$p_{23} = -1 + a_{23}(p_{23} - p_{31})$$

$$
\begin{aligned}
&= -1 + a_{23}(p_{23} - p_{21}) \quad (\text{since } p_{21} = p_{31}) \\
\text{And } \dot{p}_{21} &= -1 + a_{21}(p_{21} - p_{14}) \\
\dot{p}_{23} - \dot{p}_{21} &= a_{23}(p_{23} - p_{21}) - a_{21}(p_{21} - p_{14}) \\
&= a_{23}(p_{23} - p_{21}) - a_{21} f_{21}(t) \quad (\text{where the expression } f_{21}(t) > 0, \forall t \in [0, T)).
\end{aligned}
$$

At $t = t_{s_1}$, $(p_{23}(t_{s_1}) - p_{21}(t_{s_1})) > 0$ The solution to the above differential equation is given as,

$$
\begin{aligned}
(p_{23}(t) - p_{21}(t)) &= (p_{23}(t_{s_1}) - p_{21}(t_{s_1}))e^{a_{23}(t-t_{s_1})} - e^{a_{23}(t-t_{s_1})} \int_{t_{s_1}}^{t} e^{-a_{23}(\tau - t_{s_1})} f_{21}(\tau) d\tau \\
&= (p_{23}(t_{s_1}) - p_{21}(t_{s_1}))e^{a_{23}(t-t_{s_1})} + e^{a_{23}(t-t_{s_1})} \int_{t}^{t_{s_1}} e^{-a_{23}(\tau - t_{s_1})} f_{21}(\tau) d\tau
\end{aligned}
$$

Since $p_{23}(t_{s_1}) - p_{21}(t_{s_1}) > 0$, it can be observed that $p_{23}(t) - p_{21}(t) > 0$, $\forall t < t_{s_1}$. Similarly it can be argued that $p_{31}(t) < p_{32}(t)$, $\forall t \in [0, t_{s_1}]$ Thus the relationships $p_{21}(t) = \min \{p_{21}(t), p_{23}(t), p_{24}(t)\}$ and $p_{31}(t) = \min \{p_{31}(t), p_{32}(t), p_{34}(t)\}$, hold true $\forall t \in [0, t_{s_1}]$ The relationship $p_{14}(t) = \min \{p_{14}(t), p_{12}(t), p_{13}(t)\}$ also can be verified to hold true for all $t < t_{s_1}$ Hence the routing variables $\alpha_{14}(t)$, $\alpha_{21}(t)$ and $\alpha_{31}(t)$ continue to be unity for all $t < t_{s_1}$ Therefore regime 19 has to be the starting regime.

Thus, if the link parameters of the network are such that $a_{21} = a_{31}$ and $a_{14} > a_{24} = a_{34}$, then the optimal routing strategy for the composite four-node network is the same as synthesised from the *locally optimal* strategies for the units 1 and 3, when they are considered in isolation The switching instant can also be obtained by solving the equation,

$$
\frac{1 - e^{a_{21}(t_{s_1} - T)}}{a_{21}} + \frac{1 - e^{a_{14}(t_{s_1} - T)}}{a_{14}} + \frac{e^{a_{21}(t_{s_1} - T)} - e^{a_{14}(t_{s_1} - T)}}{(a_{21} - a_{14})} = \frac{1 - e^{a_{24}(t_{s_1} - T)}}{a_{24}}
$$

which is the same as that for the individual units when they are considered in isolation

Remark 5.4.3 : By analogous reasonings, it can be concluded that when the link parameters satisfy the following relationships,

(i) $a_{24} > a_{34} = a_{14}$

(ii) $a_{12} = a_{32}$

then optimal routing strategy for the composite network is the same as that synthesised from those for the individual units 1 and 2 Corresponding to this situation, the network starts in regime 8 and switches over to regime 27 if

$$p_{12}(0) = \frac{1 - e^{-a_{12}T}}{a_{12}} + \frac{1 - e^{-a_{24}T}}{a_{24}} + \frac{(e^{-a_{12}T} - e^{-a_{24}T})}{(a_{12} - a_{24})} < p_{14}(0) = \frac{1 - e^{-a_{14}T}}{a_{14}}$$

The switching instant t_{s_1} is obtained from solving the equation:

$$\frac{1 - e^{a_{12}(t_{s_1} - T)}}{a_{12}} + \frac{1 - e^{a_{24}(t_{s_1} - T)}}{a_{24}} + \frac{e^{a_{12}(t_{s_1} - T)} - e^{a_{24}(t_{s_1} - T)}}{(a_{12} - a_{24})} = \frac{1 - e^{a_{14}(t_{s_1} - T)}}{a_{14}}$$

Similarly if the link parameters satisfy the following conditions

(i) $a_{34} > a_{24} = a_{14}$

(ii) $a_{23} = a_{13}$

the optimal routing strategy for the network is the same as that for the units 2 and 3 when they are considered in isolation

CASE B :

For the regime 27 to be preceded by regime 21 in which $p_{21}(t) < p_{24}(t)$, it can be easily verified that the following two conditions together, are necessary and sufficient:

(i) $a_{21} > a_{31}$

(ii) $$\frac{1 - e^{(-a_{21}T)}}{a_{21}} + \frac{1 - e^{(-a_{14}T)}}{a_{14}} + \frac{e^{(-a_{21}T)} - e^{(-a_{14}T)}}{a_{21} - a_{14}} < \frac{1 - e^{(-a_{24}T)}}{a_{24}}$$

Remark 5.4.4 : Condition (i) ensures that $p_{31}(t) > p_{21}(t), \forall t \in [0, T)$, and therefore if the functions $p_{34}(t)$ and $p_{31}(t)$ intersect, then the instant at which this takes place

(denoted as t_{s_2}) is less than t_{s_1} where t_{s_1} is the instant at which $p_{21}(t)$ and $p_{24}(t)$ intersect The instant t_{s_1} is the solution to the following equation

$$\frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{(a_{21}-a_{14})} = \frac{1-e^{a_{24}(t-T)}}{a_{24}}.$$

The regime 21 might in turn, be preceded by other regimes. It can be argued along the following lines, that the only regimes that could possibly precede the regime 21 are regimes 19 and 20 In regime 21, the costate variable $p_{21}(t)$ is given as

$$\begin{aligned} p_{21}(t) &= \frac{1-e^{a_{21}(t-T)}}{a_{21}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{(a_{21}-a_{14})} \\ &= p_{14}(t) + f_{21}(t) \quad \text{where } f_{21}(t) > 0. \end{aligned}$$

The costate variable $p_{12}(t)$ satisfies the following differential equation

$$\begin{aligned} p_{12} &= -1 + a_{12}(p_{12} - p_{21}) \\ \text{Since } p_{21} > p_{14} \quad p_{12} &< -1 + a_{12}(p_{12} - p_{14}) \\ \text{where} \quad p_{14} &= -1 + a_{14}p_{14} \\ \text{Therefore} \quad p_{12} - p_{14} &< a_{12}(p_{12} - p_{14}) - a_{14}p_{14} \\ &< a_{12}(p_{12} - p_{14}) \\ \text{Let} \quad p_{12}(t) - p_{14}(t) &= a_{12}(p_{12} - p_{14}) - f_{12}(t) \quad \text{where } f_{12}(t) > 0. \\ \text{Then} \quad (p_{12}(t) - p_{14}(t)) &= (p_{12}(t_{s_1}) - p_{14}(t_{s_1}))e^{a_{12}(t-t_{s_1})} \\ &\quad - e^{a_{12}(t-t_{s_1})} \int_{t_{s_1}}^{t} e^{-a_{12}(\tau - t_{s_1})} f_{12}(\tau) d\tau \\ &= (p_{12}(t_{s_1}) - p_{14}(t_{s_1}))e^{a_{12}(t-t_{s_1})} \\ &\quad + e^{a_{12}(t-t_{s_1})} \int_{t}^{t_{s_1}} e^{-a_{12}(\tau - t_{s_1})} f_{12}(\tau) d\tau \end{aligned}$$

Since $(p_{12}(t_{s_1}) - p_{14}(t_{s_1})) > 0$ and since the second term in the above expression is positive for all $t < t_{s_1}$, it can be concluded that regime 21 can not be preceded by any regime in which $p_{12}(t) < p_{14}(t)$ (and consequently by any regime in which $\alpha_{12}(t) = 1$) Furthermore, since $a_{34} < a_{14}$ the function $p_{34}(t) > p_{14}(t)$, for all $t \in [0, T)$ and therefore the regime 21 can not also be preceded by any regime in which $\alpha_{13}(t)$

is unity. Thus in any regime that precedes the regime 21, the routing variable $\alpha_{14}(t)$ has to be unity.

Since $p_{21}(t) < p_{24}(t)$ and $p_{24}(t) < p_{23}(t)$ for all $t < t_{s_1}$, the value of the routing variable $\alpha_{21}(t)$ also can not change. Therefore the only regimes that could possibly precede the regime 21 are 19 and 20

Assume that the necessary and sufficient conditions for the regime 27 to be preceded by regime 21 are satisfied. Let us investigate the conditions under which the regime 21 is preceded by either 19 or 20

I. Link parameter $a_{32} < a_{31}$.

Under this condition, it can be argued that the costate variables $p_{32}(t)$ and $p_{31}(t)$ in regime 21 satisfy the relationship $p_{32}(t) > p_{31}(t), \forall t \in [0, t_{s_1}]$ The arguments are as given below .

In regime 21, the differential equations in the costate variable $p_{32}(t)$ and $p_{31}(t)$ are the following .

$$\begin{aligned}
p_{32} &= -1 + a_{32}(p_{32} - p_{21}) \\
\text{where } p_{21}(t) &= \frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{a_{21} - a_{14}} \\
&> p_{14}(t) = \frac{1 - e^{a_{14}(t-T)}}{a_{14}} \\
p_{31} &= -1 + a_{31}(p_{31} - p_{14}) \\
&> -1 + a_{32}(p_{31} - p_{14}) \quad \text{since } a_{32} < a_{31} \\
\text{Therefore } p_{32} - p_{31} &< a_{32}(p_{32} - p_{31}) \\
\text{Let } (p_{32} - p_{31}) &= a_{32}(p_{32} - p_{31}) - f(t) \quad \text{where } f(t) > 0, \forall t < t_{s_1}. \\
\text{Then } p_{32}(t) - p_{31}(t) &= (p_{32}(t_{s_1}) - p_{31}(t_{s_1}))e^{a_{32}(t-t_{s_1})} - e^{a_{32}t}\int_{t_{s_1}}^{t} e^{-a_{32}\tau} f(\tau)d\tau
\end{aligned}$$

$$= (p_{32}(t_{s_1}) - p_{31}(t_{s_1}))e^{a_{32}(t-t_{s_1})} + e^{a_{32}t}\int_{t}^{t_{s_1}} e^{-a_{32}\tau} f(\tau)d\tau \tag{5.62}$$

We know that

$$\begin{aligned}
p_{32}(t_{s_1}) &= \frac{1 - e^{a_{32}(t_{s_1}-T)}}{a_{32}} + \frac{1 - e^{a_{24}(t_{s_1}-T)}}{a_{24}} + \frac{e^{a_{32}(t_{s_1}-T)} - e^{a_{24}(t-T)}}{a_{32} - a_{24}} \\
p_{31}(t_{s_1}) &= \frac{1 - e^{a_{31}(t_{s_1}-T)}}{a_{31}} + \frac{1 - e^{a_{14}(t_{s_1}-T)}}{a_{14}} + \frac{e^{a_{31}(t_{s_1}-T)} - e^{a_{14}(t-T)}}{a_{31} - a_{14}}
\end{aligned}$$

Since $a_{32} < a_{31}$ and $a_{24} < a_{14}$, $p_{32}(t_{s_1}) > p_{31}(t_{s_1})$ Therefore the right hand side of the Equation (5 62) is positive $\forall t \leq t_{s_1}$ Thus regime 21 can not be preceded by any regime in which $\alpha_{32}(t) = 1$ This implies that regime 21 can not be preceded by regime 20 For the regime 21 to be preceded by regime 19, it is necessary that $p_{31}(0) < p_{34}(0)$. The switching instant t_{s_2} is given by the equation

$$\frac{1 - e^{a_{31}(t_{s_2}-T)}}{a_{31}} + \frac{1 - e^{a_{14}(t_{s_2}-T)}}{a_{14}} + \frac{e^{a_{31}(t_{s_2}-T)} - e^{a_{14}(t_{s_2}-T)}}{a_{31} - a_{14}} = \frac{1 - e^{a_{34}(t_{s_2}-T)}}{a_{34}}$$

Since the relationships

(i) $p_{31}(t) = min\{p_{31}(t), p_{32}(t), p_{34}(t)\}$

(ii) $p_{21}(t) = min\{p_{21}(t), p_{23}(t), p_{24}(t)\}$

(iii) $p_{14}(t) = min\{p_{14}(t), p_{13}(t), p_{12}(t)\}$

hold true $\forall t < t_{s_2}$, the regime 19 must be the starting regime itself. Thus it can be observed that under this situation also, the optimal routing strategy for the network can be synthesised from those for the individual units 1 and 3 when they are considered in isolation. The inter regime transition for this case is shown in the Figure 5.18

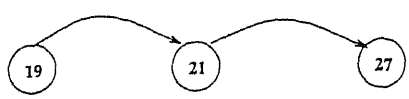

Figure 5 18. Inter regime transition diagram for the case $a_{14} > a_{24} = a_{34}$, $a_{32} < a_{31}$

II. Link parameter $a_{32} > a_{31}$ **:**

Let us examine all the possibilities under these conditions As mentioned earlier, let t_{s_1} denote the inter regime transition instant between regimes 21 and 27 To study the various possibilities, we focus our attention on the following conditions (a) and (b) given below

Condition (a) : $p_{32}(t_{s_1}) < p_{31}(t_{s_1})$

i e. the link parameters satisfy the inequality

$$\left.\begin{matrix} \frac{1-e^{a_{32}(t_{s_1}-T)}}{a_{32}} + \frac{1-e^{a_{24}(t_{s_1}-T)}}{a_{24}} \\ +\frac{e^{a_{32}(t_{s_1}-T)}-e^{a_{24}(t_{s_1}-T)}}{a_{32}-a_{24}} \end{matrix}\right\} < \left\{\begin{matrix} \frac{1-e^{a_{31}(t_{s_1}-T)}}{a_{31}} + \frac{1-e^{a_{14}(t_{s_1}-T)}}{a_{14}} \\ +\frac{e^{a_{31}(t_{s_1}-T)}-e^{a_{14}(t_{s_1}-T)}}{a_{31}-a_{14}} \end{matrix}\right. \tag{5 63}$$

Condition (b) : $p_{32}(0) < p_{31}(0)$

where $p_{32}(t)$ is the solution to the differential equation $p_{32} = -1 + a_{32}(p_{32} - p_{21})$ with the initial condition

$$p_{32}(t_{s_1}) = \frac{1 - e^{a_{32}(t_{s_1}-T)}}{a_{32}} + \frac{1 - e^{a_{21}(t_{s_1}-T)}}{a_{21}} + \frac{e^{a_{32}(t_{s_1}-T)} - e^{a_{24}(t_{s_1}-T)}}{a_{32} - a_{24}}$$

and $p_{21}(t)$ is given by

$$p_{21}(t) = \frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{a_{21} - a_{14}}$$

The solution $p_{32}(t)$ of the above differential equation is given as

$$\begin{aligned} p_{32}(t) = {} & p_{32}(t_{s_1})e^{a_{32}(t-t_{s_1})} + (1 - e^{a_{32}(t-t_{s_1})})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ & + \frac{a_{14}a_{32}e^{a_{21}(t_{s_1}-T)}}{a_{21}(a_{14} - a_{21})(a_{21} - a_{32})}(e^{a_{21}(t-t_{s_1})} - e^{a_{32}(t-t_{s_1})}) \\ & + \frac{a_{21}a_{32}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(t-t_{s_1})} - e^{a_{32}(t-t_{s_1})})}{a_{14}(a_{21} - a_{14})(a_{14} - a_{32})} \end{aligned} \tag{5.64}$$

and therefore $p_{32}(0)$ in the condition (b) is given by the following expression

$$\begin{aligned} p_{32}(0) = {} & p_{32}(t_{s_1})e^{-a_{32}t_{s_1}} + (1 - e^{-a_{32}t_{s_1}})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ & + \frac{a_{14}a_{32}e^{a_{21}(t_{s_1}-T)}}{a_{21}(a_{14} - a_{21})(a_{21} - a_{32})}(e^{-a_{21}t_{s_1}} - e^{-a_{32}t_{s_1}}) \\ & + \frac{a_{21}a_{32}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(-t_{s_1})} - e^{a_{32}(-t_{s_1})})}{a_{14}(a_{21} - a_{14})(a_{14} - a_{32})} \end{aligned} \tag{5 65}$$

If both the conditions (a) and (b) given above are satisfied, then the function $p_{31}(t) > p_{32}(t)$, $\forall t \in [0, t_{s_1}]$. This in turn implies that the regime 21 can not be preceded by regime 19 in which the routing variable $\alpha_{31}(t) = 1$ If $p_{32}(0)$ as given by the above expression is such that

$$p_{32}(0) < p_{34}(0) = \frac{1 - e^{-a_{34}T}}{a_{34}}$$

then the network operation starts in the regime 20 and switches over to regime 21 at t_{s_2} where the switching instant is obtained from solving the equation $p_{32}(t) = p_{34}(t)$. If $p_{32}(0) > p_{34}(0)$, then it can easily be inferred that the network operation must have started in the regime 21 itself.

Let condition (a) be satisfied but (b) is violated. i e $p_{32}(t_{s_1}) < p_{31}(t_{s_1})$ but $p_{32}(0) > p_{31}(0)$ The following four cases 1, 2 and 3 given below exhaust all the possibilities

Case 1 : $p_{32}(0) > p_{31}(0) > p_{34}(0)$

It can be easily inferred that regime 21 must be the starting regime

Case 2 : $p_{31}(0) < p_{34}(0)$ and $p_{32}(0) > p_{34}(0)$

It can be easily argued that regime 21 is preceded by regime 19 which is the starting regime The switching instant t_{s_2} is obtained by solving the equation $p_{31}(t) = p_{34}(t)$.

Case 3 : $p_{31}(0) < p_{34}(0)$ and $p_{32}(0) < p_{34}(0)$.

Find the switching instant between $p_{31}(t)$ and $p_{32}(t)$ where $p_{32}(t)$ is given by the Equation (5 64) Let this be denoted by t_{12} Similarly let the switching instant between $p_{31}(t)$ and $p_{34}(t)$ be denoted by t_{14} and that between $p_{32}(t)$ and $p_{34}(t)$ be denoted by t_{24}.

If $t_{24} > t_{12}$, then over $[t_{12}, t_{24}]$ the network operates in regime 20, and this is preceded by regime 19 (which is the starting regime). The inter regime transition diagram corresponding to this situation is shown in Figure 5 19.

If $t_{24} < t_{12}$, then during the interval $[0, t_{14}]$ the network operates in regime 19

Figure 5 19 Regime transition diagram for the case $t_{24} > t_{12}$

and over the interval $[t_{14}, t_{s_1}]$ the network operates in regime 21. The inter regime transition diagram is shown in Figure 5.20

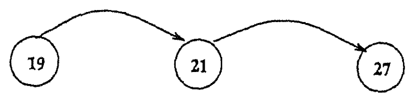

Figure 5.20: Regime transition diagram for the case $t_{24} < t_{12}$

Let condition (a) not hold true but (b) is satisfied. i e. $p_{32}(t_{s_1}) > p_{31}(t_{s_1})$ and $p_{32}(0) < p_{31}(0)$. This implies that the functions $p_{32}(t)$ and $p_{31}(t)$ intersect in the interval $[0, t_{s_1}]$ Let t_{12} denote this instant Then $p_{32}(t) < p_{31}(t)$ in $[0, t_{12}]$ and $p_{32}(t) > p_{31}(t)$ in $[t_{12}, t_{s_1}]$ If $p_{32}(0)$ is greater than $p_{34}(0)$ then both the functions $p_{32}(t)$ and $p_{31}(t)$ are greater than $p_{34}(t)$ in $[0, t_{s_1}]$ Therefore regime 21 must have been the starting regime If $p_{32}(0)$ is less than $p_{34}(0)$ then the following two possibilities arise:

(i) $p_{31}(0) > p_{34}(0)$. This implies that the function $p_{31}(t)$ is greater than $p_{34}(t)$ in the interval $[0, t_{s_1}]$ while $p_{32}(t)$ and $p_{34}(t)$ intersect Let us denote this instant as t_{24}. The network operates in regime 20 during $[0, t_{24})$, in regime 21 during $[t_{24}, t_{s_1})$ and in regime 27 during $[t_{s_1}, T]$ as shown in Figure 5 21.

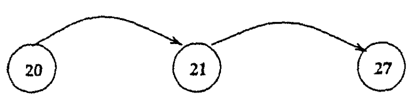

Figure 5 21. The regime transition diagram for case (i)

(ii) $p_{31}(0) < p_{34}(0)$ The routing strategy can be specified as follows
Find the instant when the functions $p_{31}(t)$ and $p_{32}(t)$ intersect Let this be denoted as denoted as t_{12} The instant of intersection of $p_{31}(t)$ and $p_{34}(t$ be denoted as t_{14} If $t_{12} < t_{24}$, (where t_{24} is the instant when $p_{32}(t)$ and $p_{34}(t)$ intersect) then the regime 21 is preceded by regime 20 which is further preceded by regime 19 The

network operates in regime 19 during $[0, t_{12}]$, in regime 20 during $(t_{12}, t_{24}]$, in regime 21 during $(t_{24}, t_{s_1}]$ and in regime 27 (terminal regime) during $(t_{s_1}, T]$ as shown in Figure 5 22

Figure 5 22: Regime transition diagram for the case $t_{12} < t_{24}$

If on the other hand, $t_{12} > t_{24}$, then the regime 21 is preceded by regime 19 which in turn is preceded by regime 20. The network operates in regime 20 during $[0, t_{12}]$, in regime 19 during $(t_{12}, t_{14}]$, in regime 21 during $(t_{14}, t_{s_1}]$ and in regime 27 (terminal regime) during $(t_{s_1}, T]$ The regime transition diagram is shown in Figure 5 23

Figure 5.23 Regime transition diagram for the case $t_{12} > t_{24}$

If both the conditions (a) and (b) are violated then it can be seen that regime 21 can not be preceded by regime 20. Either the network operation has started in regime 21 or in regime 19 The necessary and sufficient condition under which the network operation starts in regime 19 is given as

$$\frac{1-e^{-a_{31}T}}{a_{31}} + \frac{1-e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{31}T}-e^{-a_{14}T}}{a_{31}-a_{14}} < \frac{1-e^{-a_{34}T}}{a_{34}}$$

The switching instant between regime 19 and regime 21 is obtained by solving the following equation

$$\frac{1-e^{a_{31}(t-T)}}{a_{31}} + \frac{1-e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{31}(t-T)}-e^{a_{14}(t-T)}}{a_{31}-a_{14}} < \frac{1-e^{a_{34}(t-T)}}{a_{34}}$$

The above discussion exhausts all the possible regime transitions and specifies the

equations for the inter regime transition instants and therefore the optimal routing strategy for CASE B.

CASE C :

Because of the topological symmetry with the situation in CASE B, the arguments used there can be easily extended to this case also. This being the case, the arguments are not repeated in this section. Thus the necessary and sufficient conditions for the regime 27 to be preceded by regime 25 are the following :

(i) $a_{31} > a_{21}$.

(ii) $$\frac{1 - e^{(-a_{31}T)}}{a_{31}} + \frac{1 - e^{(-a_{14}T)}}{a_{14}} + \frac{e^{(-a_{31}T)} - e^{(-a_{14}T)}}{a_{31} - a_{14}} < \frac{1 - e^{(-a_{34}T)}}{a_{34}}$$

Condition (i) above ensures that $p_{21}(t) > p_{31}(t), \forall t \in [0, T)$ and therefore the intersection between $p_{21}(t)$ and $p_{24}(t)$ (if any) takes place at an instant which is less than the instant at which $p_{31}(t)$ and $p_{34}(t)$ intersect Furthermore, it can be argued that regime 25 can possibly be preceded by regimes 19 and 22 only in which the routing variable $\alpha_{31}(t)$ is unity If the link parameters are such that $a_{23} < a_{21}$ then regime 25 can not be preceded by regime 22 Corresponding to this situation, either the regime 25 is the starting regime or is preceded by regime 19

The additional necessary and sufficient condition for regime 25 to be preceded by regime 19 is given as

$$\frac{1 - e^{-a_{21}T}}{a_{21}} + \frac{1 - e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{21}T} - e^{-a_{14}T}}{a_{21} - a_{14}} < \frac{1 - e^{-a_{24}T}}{a_{24}}$$

And the switching instant t_{s_2} between regime 19 and regime 25 is obtained by solving the equation

$$\frac{1 - e^{a_{21}(t_{s_2} - T)}}{a_{21}} + \frac{1 - e^{a_{14}(t_{s_2} - T)}}{a_{14}} + \frac{e^{a_{21}(t_{s_2} - T)} - e^{a_{14}(t_{s_2} - T)}}{a_{21} - a_{14}} = \frac{1 - e^{a_{24}(t_{s_2} - T)}}{a_{24}}$$

The regime transition diagram for this case is shown in Figure 5 24

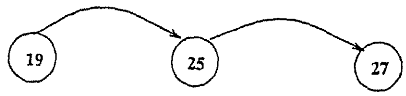

Figure 5 24. Regime transition diagram for the case $a_{23} < a_{21}$

When $a_{23} > a_{21}$, consider the following two conditions :

Condition (a) : $p_{23}(t_{s_1}) < p_{21}(t_{s_1})$ i.e. the link parameters satisfy the inequality

$$\left.\begin{array}{c} \frac{1-e^{a_{23}(t_{s_1}-T)}}{a_{23}} + \frac{1-e^{a_{34}(t_{s_1}-T)}}{a_{34}} \\ +\frac{e^{a_{23}(t_{s_1}-T)}-e^{a_{34}(t_{s_1}-T)}}{a_{23}-a_{34}} \end{array}\right\} < \left\{\begin{array}{c} \frac{1-e^{a_{21}(t_{s_1}-T)}}{a_{21}} + \frac{1-e^{a_{14}(t_{s_1}-T)}}{a_{14}} \\ +\frac{e^{a_{21}(t_{s_1}-T)}-e^{a_{14}(t_{s_1}-T)}}{a_{21}-a_{14}} \end{array}\right. \qquad (5\ 66)$$

Condition (b) : $p_{23}(0) < p_{21}(0)$ where $p_{23}(t)$ is given as

$$\begin{aligned} p_{23}(t) &= p_{23}(t_{s_1})e^{a_{23}(t-t_{s_1})} + (1-e^{a_{23}(t-t_{s_1})})(1/a_{23}+1/a_{31}+1/a_{14}) \\ &+\frac{a_{14}a_{23}e^{a_{31}(t_{s_1}-T)}}{a_{31}(a_{14}-a_{31})(a_{31}-a_{23})}(e^{a_{31}(t-t_{s_1})}-e^{a_{23}(t-t_{s_1})}) \\ &+\frac{a_{31}a_{23}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(t-t_{s_1})}-e^{a_{23}(t-t_{s_1})})}{a_{14}(a_{31}-a_{14})(a_{14}-a_{23})} \end{aligned} \qquad (5.67)$$

and therefore $p_{23}(0)$ in the condition (b) above is given by the expression.

$$\begin{aligned} p_{23}(0) &= p_{23}(t_{s_1})e^{a_{23}(-t_{s_1})} + (1-e^{a_{23}(-t_{s_1})})(1/a_{23}+1/a_{31}+1/a_{14}) \\ &+\frac{a_{14}a_{23}e^{a_{31}(t_{s_1}-T)}}{a_{31}(a_{14}-a_{31})(a_{31}-a_{23})}(e^{a_{31}(-t_{s_1})}-e^{a_{23}(-t_{s_1})}) \\ &+\frac{a_{31}a_{23}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(-t_{s_1})}-e^{a_{23}(-t_{s_1})})}{a_{14}(a_{31}-a_{14})(a_{14}-a_{23})} \end{aligned} \qquad (5.68)$$

If both the conditions (a) and (b) given above are satisfied, then the regime 25 can not be preceded by regime 19 If $p_{23}(0)$ as given by the above expression is such that

$$p_{23}(0) \ < \ p_{24}(0) = \frac{1-e^{-a_{24}T}}{a_{24}}$$

then the network operation starts in regime 22 and switches over to regime 25 at t_{s_2} where the switching instant t_{s_2} is obtained from solving the equation $p_{23}(t) = p_{24}(t)$. If $p_{23}(0) > p_{24}(0)$, then the network operation must have started in regime 25 itself.

Let condition (a) be satisfied but (b) is violated. Then the following three cases exhaust all the possibilities :

Case 1 : $p_{23}(0) > p_{21}(0) > p_{24}(0)$.

It can be easily inferred that regime 25 is the starting regime

Case 2 : $p_{21}(0) < p_{24}(0)$ and $p_{23}(0) > p_{24}(0)$

It can be easily argued that regime 25 is preceded by regime 19 which is the starting regime. The switching instant t_{s_2} can be obtained by solving the equation $p_{21}(t) = p_{24}(t)$.

Case 3 : $p_{21}(0) < p_{24}(0)$ and $p_{23}(0) < p_{24}(0)$ Corresponding to this situation the optimal routing strategy can be specified as follows

Find the switching instant between $p_{21}(t)$ and $p_{23}(t)$ where $p_{23}(t)$ is given by the Equation (5.68). Let this be denoted by t_{13} Similarly let the switching instant between $p_{21}(t)$ and $p_{24}(t)$ be denoted by t_{14} and that between $p_{23}(t)$ and $p_{24}(t)$ be denoted by t_{34}. If $t_{34} > t_{13}$, then during the interval $[t_{13}, t_{34}]$ the network operates in regime 22, and this is preceded by regime 19 (which is the starting regime) The regime transition diagram is shown in Figure 5.25.

Figure 5.25. Regime transition diagram for the case $t_{34} > t_{13}$

If $t_{34} < t_{13}$, then during the interval $[0, t_{14}]$, the network operates in regime 19 and over the interval $[t_{14}, t_{s_1}]$ the network operates in regime 25. The regime transition diagram for this case is shown in Figure 5.26.

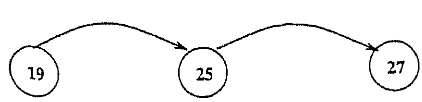

Figure 5.26 Regime transition diagram for the case $t_{34} < t_{13}$

Let condition (a) not hold true but (b) holds true i e $p_{23}(t_{s_1}) > p_{21}(t_{s_1})$ and $p_{23}(0) < p_{21}(0)$ This implies that $p_{23}(t)$ and $p_{21}(t)$ intersect in the interval $[0, t_{s_1}]$ Let t_{13} denote this instant Then $p_{23}(t) < p_{21}(t)$ in $[0, t_{13}]$ and $p_{23}(t) > p_{21}(t)$ in $[t_{13}, t_{s_1}]$ If $p_{23}(0)$ is greater than $p_{24}(0)$ then both the functions $p_{23}(t)$ and $p_{21}(t)$ are greater than $p_{24}(t)$ in $[0, t_{s_1}]$ Therefore regime 25 must have been the starting regime If $p_{23}(0)$ is less than $p_{24}(0)$ then the following two possibilities arise

(i) $p_{21}(0) > p_{24}(0)$ This implies that the function $p_{21}(t)$ is greater than $p_{24}(t)$ in the interval $[0, t_{s_1}]$ while $p_{23}(t)$ and $p_{24}(t)$ intersect. Let us denote this instant as t_{34} The network operates in regime 22 during $[0, t_{34})$. During the interval $[t_{34}, t_{s_1})$ the network operates in regime 25 and during $[t_{s_1}, T]$ the network operates in regime 27. The regime transition diagram is shown in Figure 5.27.

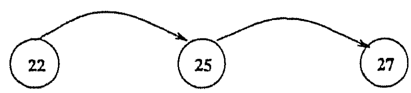

Figure 5.27· The regime transition diagram for case (i)

(ii) $p_{21}(0) < p_{24}(0)$ The routing strategy can be specified as follows

Find the instant when the functions $p_{21}(t)$ and $p_{23}(t)$ intersect Let this be denoted as denoted as t_{13}. The instant of intersection of $p_{21}(t)$ and $p_{24}(t)$ be denoted as t_{14} If $t_{13} < t_{34}$, (where t_{34} is the instant when $p_{23}(t)$ and $p_{24}(t)$ intersect) then the regime 25 is preceded by regime 22 which is further preceded by regime 19 The network operates in regime 19 during $[0, t_{13}]$, in regime 22 during $(t_{13}, t_{34}]$, in regime 25 during $(t_{34}, t_{s_1}]$ and in regime 27 (terminal regime) during $(t_{s_1}, T]$. The regime transition diagram is shown in Figure 5.28 If on the other hand, $t_{13} > t_{34}$, then

Figure 5.28· Regime transition diagram for the case $t_{13} < t_{34}$

the regime 25 is preceded by regime 19 which in turn is preceded by regime 22

Figure 5.29: Regime transition diagram for the case $t_{13} > t_{34}$

The network operates in regime 22 during $[0, t_{13}]$, in regime 19 during $(t_{13}, t_{14}]$, in regime 25 during $(t_{14}, t_{s_1}]$ and in regime 27 (terminal regime) during $(t_{s_1}, T]$. The regime transition diagram is shown in Figure 5.29.

If both the conditions (a) and (b) are violated then it can be seen that regime 25 can not be preceded by regime 22. Either the network operation has started in regime 25 or in regime 19. The necessary and sufficient condition under which the network operation starts in regime 19 is given as

$$\frac{1 - e^{-a_{21}T}}{a_{21}} + \frac{1 - e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{21}T} - e^{-a_{14}T}}{a_{21} - a_{14}} < \frac{1 - e^{-a_{24}T}}{a_{24}}.$$

The switching instant between regime 19 and regime 25 is obtained by solving the following equation :

$$\frac{1 - e^{a_{21}(t-T)}}{a_{21}} + \frac{1 - e^{a_{14}(t-T)}}{a_{14}} + \frac{e^{a_{21}(t-T)} - e^{a_{14}(t-T)}}{a_{21} - a_{14}} < \frac{1 - e^{a_{24}(t-T)}}{a_{24}}.$$

The above three **CASES A, B and C** exhaust all the possible regime transitions when link (1,4) is the *fastest* direct link and the links (2,4) and (3,4) are of the same parameter. From the above discussion, the following two conclusions can be drawn for this case $a_{14} > a_{24} = a_{34}$:

- Traffic arriving at node 1 is routed onto the direct link (1,4) for the entire duration of network operation.

- The network operation never enters regime 23, 24 and 26. This implies that traffic arriving at node 2 is never routed through the set of links $\{(2,3),(3,4)\}$ to the destination. Similarly traffic arriving at node 3 is never sent on the set of links $\{(3,2),(2,4)\}$ to the destination. This is in accordance with our intuition since the links (2,4) and (3,4) are of the same link parameters.

The arguments employed in the above discussion to obtain the various inter regime transition instants can be extended to the two other cases and it can be concluded that when link (2,4) is the *fastest* direct link and links (1,4) and (3,4) are of the same parameter, then the link (2,4) is used for the entire duration of network operation. For this choice of the link parameters the network operation never enters regimes 16, 18 and 25. Similarly when link (3,4) is the *fastest* direct link with links (1,4) and (2,4) of the same link parameter, then the network operation never enters regimes 3, 9 and 21. Link (3,4) is used for the entire duration of network operation.

5.4.2 Link (1,4) slower than (2,4) and (3,4)

Consider the situation wherein the link (1,4) is *slower* than links (2,4) and (3,4) (which have the same parameter). i.e. $a_{14} < a_{24} = a_{34}$.
It is easily verified that under this situation the following relationships hold true:

$$\begin{aligned}
p_{14}(t) = \frac{1 - e^{a_{14}(t-T)}}{a_{14}} &> p_{24}(t) = \frac{1 - e^{a_{24}(t-T)}}{a_{24}} \quad \forall t \in [0, T).\\
p_{24}(t) &< p_{21}(t) \quad \forall t \in [0, T).\\
p_{24}(t) &< p_{23}(t) \quad \forall t \in [0, T).\\
p_{34}(t) &< p_{31}(t) \quad \forall t \in [0, T).\\
p_{34}(t) &< p_{32}(t) \quad \forall t \in [0, T).
\end{aligned}$$

Therefore the terminal regime 27 can not be preceded by any regime in which $\alpha_{24}(t)$ and/or $\alpha_{34}(t)$ is zero. Thus the regimes which could possibly precede the regime 27 are 9 and 18 only. In regime 9, the routing variable $\alpha_{12}(t)$ is unity. A necessary condition for regime 27 to be preceded by regime 9 is that $a_{12} > a_{13}$. This condition ensures that $p_{12}(t) < p_{13}(t)$ in the interval $[0.T)$ and therefore $\alpha_{13}(t)$ is zero for the entire interval of operation. The following (additional) condition is sufficient for

the network operation to start in regime 9 and switch over to regime 27 :

$$p_{12}(0) = \frac{1-e^{-a_{12}T}}{a_{12}} + \frac{1-e^{-a_{24}T}}{a_{24}} + \frac{e^{-a_{12}T}-e^{-a_{24}T}}{a_{12}-a_{24}} < p_{14}(0) = \frac{1-e^{-a_{14}T}}{a_{14}}.$$

A necessary condition for regime 27 to be preceded by 18 is $a_{13} > a_{12}$. The following (additional) condition is sufficient for the network operation to start in regime 18 and switch over to regime 27.

$$p_{13}(0) = \frac{1-e^{-a_{13}T}}{a_{13}} + \frac{1-e^{-a_{34}T}}{a_{34}} + \frac{e^{-a_{13}T}-e^{-a_{34}T}}{a_{13}-a_{34}} < p_{14}(0) = \frac{1-e^{-a_{14}T}}{a_{14}}.$$

By analogous arguments, when $a_{14} = a_{24} > a_{34}$ the network operation can possibly start either in regime 25 or 26 only. The necessary and sufficient conditions for the network operation to start in regime 25 are

(i) $a_{31} > a_{32}$

(ii) $$\frac{1-e^{(-a_{31}T)}}{a_{31}} + \frac{1-e^{(-a_{14}T)}}{a_{14}} + \frac{e^{(-a_{31}T)}-e^{(-a_{14}T)}}{a_{31}-a_{14}} < \frac{1-e^{(-a_{34}T)}}{a_{34}}.$$

The necessary and sufficient conditions for the network operation to start in regime 26 are

(i) $a_{32} > a_{31}$

(ii) $$\frac{1-e^{(-a_{32}T)}}{a_{32}} + \frac{1-e^{(-a_{24}T)}}{a_{24}} + \frac{e^{(-a_{32}T)}-e^{(-a_{24}T)}}{a_{32}-a_{24}} < \frac{1-e^{(-a_{34}T)}}{a_{34}}.$$

For the case wherein $a_{14} = a_{34} > a_{24}$, the terminal regime 27 can possibly be preceded by regimes 21 and 24 only. The necessary and sufficient conditions for the network operation to start in regime 21 are

(i) $a_{21} > a_{23}$

(ii) $$\frac{1-e^{(-a_{21}T)}}{a_{21}} + \frac{1-e^{(-a_{14}T)}}{a_{14}} + \frac{e^{(-a_{21}T)}-e^{(-a_{14}T)}}{a_{21}-a_{14}} < \frac{1-e^{(-a_{24}T)}}{a_{24}}.$$

The necessary and sufficient conditions for the network operation to start in regime 24 are

(i) $a_{23} > a_{21}$

(ii) $$\frac{1-e^{(-a_{23}T)}}{a_{23}} + \frac{1-e^{(-a_{34}T)}}{a_{34}} + \frac{e^{(-a_{23}T)} - e^{(-a_{34}T)}}{a_{23}-a_{34}} < \frac{1-e^{(-a_{24}T)}}{a_{24}}.$$

From the above discussion the following conclusions can be drawn regarding the optimal routing strategy for the case wherein link (1,4) is *slower* than links (2,4) and (3,4) (which have the same link parameter) :

- The traffic arriving at the nodes 2 and 3 are routed onto the direct links to the destination for the entire duration of network operation.
- Traffic arriving at node 1 is never routed through the set of links $\{(1,2),(2.4)\}$ if the link (1,2) is *slower* than link (1,3).
- The traffic at node 1 is never routed through the set of links $\{(1,3),(3.4)\}$ if the link (1,3) is *slower* than link (1,2).

5.5 The direct links with different parameters

Consider the case where the direct links have different link parameters. Let us assume that $a_{14} > a_{24} > a_{34}$. Because of the topological symmetry of the network the arguments in this section applies to the other cases in an exactly similar manner. Let us investigate the possible regime transitions and the equations to obtain the switching instants.

Since $a_{14} > a_{24} > a_{34}$, by Theorem 5.4.3, the terminal regime 27 can not be preceded by regimes 1 to 18. Furthermore since $a_{24} > a_{34}$

$$\begin{aligned} p_{24}(t) &< p_{34}(t) \quad \forall t \in [0,T) \\ p_{23}(t) &> p_{24}(t) \quad \forall t \in [0,T) \end{aligned}$$

Therefore regime 27 can not be preceded by regimes 22, 23 and 24 in which $\alpha_{23}(t)$ is unity. Thus the regimes that could possibly precede the regime 27 are regimes 19, 20, 21, 25 and 26 only. Let us consider each of these possibilities in detail.

In regime 19, the routing variables $\alpha_{21}(t)$ and $\alpha_{31}(t)$ are unity. For the regime 27 to be preceded by regime 19, it is necessary that the functions $p_{21}(t)$ and $p_{24}(t)$ as given by the Equations (5.53) and (5.55) intersect at some t_{s_1}. It is also necessary that the functions $p_{31}(t)$ and $p_{34}(t)$ as given by the Equations (5.56) and (5.58) intersect at the same instant t_{s_1}. Furthermore at $t = t_{s_1}$, the function $p_{32}(t)$ has to be greater than $p_{34}(t)$ so that regime 27 is not preceded by regime 20. Thus the set of necessary and sufficient conditions under which the regime 27 is preceded by regime 19 are given below :

(i) $p_{21}(0) < p_{24}(0)$

(ii) $p_{31}(0) < p_{34}(0)$

(iii) Let t_{s_1} be the instant at which $p_{21}(t) = p_{24}(t)$. Then $p_{31}(t_{s_1}) = p_{34}(t_{s_1})$ and $p_{32}(t_{s_1}) > p_{34}(t_{s_1})$.

The above conditions when stated in terms of the link parameters are as follows:

$$\frac{1-e^{-a_{21}T}}{a_{21}} + \frac{1-e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{21}T} - e^{-a_{14}T}}{a_{21} - a_{14}} < \frac{1-e^{-a_{24}T}}{a_{24}} \tag{5.69}$$

$$\frac{1-e^{-a_{31}T}}{a_{31}} + \frac{1-e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{31}T} - e^{-a_{14}T}}{a_{31} - a_{14}} < \frac{1-e^{-a_{34}T}}{a_{34}} \tag{5.70}$$

$$\frac{1-e^{a_{31}(t_{s_1}-T)}}{a_{31}} + \frac{1-e^{a_{14}(t_{s_1}-T)}}{a_{14}} + \frac{e^{a_{31}(t_{s_1}-T)} - e^{a_{14}(t_{s_1}-T)}}{a_{31} - a_{14}} = \frac{1-e^{a_{34}(t_{s_1}-T)}}{a_{34}} \tag{5.71}$$

$$\frac{1-e^{a_{32}(t_{s_1}-T)}}{a_{32}} + \frac{1-e^{a_{24}(t_{s_1}-T)}}{a_{24}} + \frac{e^{a_{32}(t_{s_1}-T)} - e^{a_{24}(t_{s_1}-T)}}{a_{32} - a_{24}} > \frac{1-e^{a_{34}(t_{s_1}-T)}}{a_{34}} \tag{5.72}$$

where t_{s_1} is obtained by solving the equation

$$\frac{1-e^{a_{21}(t_{s_1}-T)}}{a_{21}} + \frac{1-e^{a_{14}(t_{s_1}-T)}}{a_{14}} + \frac{e^{a_{21}(t_{s_1}-T)} - e^{a_{14}(t_{s_1}-T)}}{a_{21} - a_{14}} = \frac{1-e^{a_{24}(t_{s_1}-T)}}{a_{24}} \tag{5.73}$$

It can be verified that the relationships $p_{14}(t) = \min\{p_{12}(t), p_{13}(t), p_{14}(t)\}$ and $p_{21}(t) = \min\{p_{21}(t), p_{23}(t), p_{24}(t)\}$ hold true $\forall t < t_{s_1}$ and therefore regime 19 can not be

preceded by any regime in which $\alpha_{14}(t)$ and/or $\alpha_{21}(t)$ is zero. Thus regime 19 can possibly be preceded by regime 20 only. For the condition (5.72) above to be satisfied, it is sufficient though not necessary that $a_{31} > a_{32}$ (this along with the relationship $a_{14} > a_{24}$ ensures that $p_{32}(t) > p_{31}(t)$, $\forall t \in [0, T)$). If this is satisfied then the regime 19 can not be preceded by regime 20 and therefore has to be starting regime. If on the other hand, this condition is violated then a necessary and sufficient condition for regime 19 to be preceded by regime 20 is $p_{32}(0) < p_{31}(0)$. The equation for $p_{32}(t)$ in regime 19 is given as

$$\begin{aligned} p_{32}(t) &= p_{32}(t_{s_1})e^{a_{32}(t-t_{s_1})} + (1 - e^{a_{32}(t-t_{s_1})})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ &+ \frac{a_{14}a_{32}e^{a_{21}(t_{s_1}-T)}}{a_{21}(a_{14} - a_{21})(a_{21} - a_{32})}(e^{a_{21}(t-t_{s_1})} - e^{a_{32}(t-t_{s_1})}) \\ &+ \frac{a_{21}a_{32}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(t-t_{s_1})} - e^{a_{32}(t-t_{s_1})})}{a_{14}(a_{21} - a_{14})(a_{14} - a_{32})} \end{aligned} \tag{5.74}$$

and therefore $p_{32}(0)$ is given by the following expression :

$$\begin{aligned} p_{32}(0) &= p_{32}(t_{s_1})e^{-a_{32}t_{s_1}} + (1 - e^{-a_{32}t_{s_1}})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ &+ \frac{a_{14}a_{32}e^{a_{21}(t_{s_1}-T)}}{a_{21}(a_{14} - a_{21})(a_{21} - a_{32})}(e^{-a_{21}t_{s_1}} - e^{-a_{32}t_{s_1}}) \\ &+ \frac{a_{21}a_{32}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(-t_{s_1})} - e^{a_{32}(-t_{s_1})})}{a_{14}(a_{21} - a_{14})(a_{14} - a_{32})}. \end{aligned} \tag{5.75}$$

The instant of switching t_{s_2} between the starting regime 20 and the intermediate regime 19 is obtained by solving the equation $p_{32}(t_{s_2}) = p_{31}(t_{s_2})$ where $p_{32}(t)$ and $p_{31}(t)$ are given by Equations (5.74) and (5.56) respectively.

The conditions under which the regime 27 is preceded by regime 20 can also be easily argued out to be the following:

(i) $p_{21}(0) < p_{24}(0)$

(ii) $p_{32}(0) < p_{34}(0)$

(iii) Let t_{s_1} be the instant at which $p_{21}(t) = p_{24}(t)$. Then $p_{32}(t_{s_1}) = p_{34}(t_{s_1})$ and $p_{31}(t_{s_1}) > p_{34}(t_{s_1})$.

It can also be inferred that regime 20 can possibly be preceded by regime 19 only. The additional necessary and sufficient condition under which the network operation starts in regime 19 is that $p_{31}(0) < p_{32}(0)$ where the expression for $p_{32}(0)$ is the same as in Equation (5.75). The switching instant t_{s_2} between the starting regime 19 and intermediate regime 20 is obtained by solving the equation $p_{32}(t_{s_2}) = p_{31}(t_{s_2})$ where $p_{32}(t)$ is given by Equation (5.74) and $p_{31}(t)$ is given by the Equation (5.56). The conditions under which the terminal regime 27 is preceded by regime 21 can be stated as follows:

(i) $p_{21}(0) < p_{24}(0)$

(ii) Let t_{s_1} be the switching instant between $p_{21}(t)$ and $p_{24}(t)$. Then $p_{31}(t_{s_1}) > p_{34}(t_{s_1})$ and $p_{32}(t_{s_1}) > p_{34}(t_{s_1})$.

It can be verified that the relationships $p_{14}(t)$=min $\{p_{12}(t), p_{13}(t), p_{14}(t)\}$ and $p_{21}(t)$ = min $\{p_{21}(t), p_{23}(t), p_{24}(t)\}$, hold true $\forall t < t_{s_1}$. Therefore regime 21 can not be preceded by any regime in which $\alpha_{14}(t)$ and/or $\alpha_{21}(t)$ is zero. Thus regime 21 can be possibly be preceded by regime 19 and 20 only. The conditions under which regime 21 is preceded by 19 are the following :

(i) $p_{31}(0) < p_{34}(0)$, where $p_{31}(t)$ and $p_{34}(t)$ are given by the Equations (5.56) and (5.58) respectively.

(ii) Let the switching instant between the functions $p_{31}(t)$ and $p_{34}(t)$ be at $t = t_{s_2}$. The value of $p_{32}(t_{s_2})$ has to be greater than $p_{34}(t_{s_2})$.

The second condition above ensures that regime 21 is not preceded by regime 20. The regime 19 can in turn be preceded by regime 20. The necessary and sufficient condition for this to take place is $p_{32}(0) < p_{31}(0)$. Since the functions $p_{32}(t)$ and $p_{31}(t)$ can have at the most one point of intersection in $[0, t_{s_2}]$, the regime 20 has to be the starting regime.

The conditions under which the regime 21 (that precedes the regime 27) is preceded by regime 20 can be stated as follows:

(i) $p_{32}(0) < p_{34}(0)$ where $p_{32}(t)$ and $p_{34}(t)$ are given by the Equations (5.74) and (5.58) respectively.

(ii) Let the switching instant between the functions $p_{32}(t)$ and $p_{34}(t)$ be at $t = t_{s_2}$. Then $p_{31}(t_{s_2})$ has to be greater than $p_{34}(t_{s_2})$.

The regime 20 will be preceded by regime 19 if $p_{31}(0) < p_{32}(0)$. The instant of switching is given by the solution to the equation $p_{32}(t) = p_{31}(t)$.
The necessary and sufficient conditions under which the regime 27 is preceded by regime 25 are the following :

(i) $p_{31}(0) < p_{34}(0)$

(ii) Let the instant at which the functions $p_{31}(t)$ and $p_{34}(t)$ intersect be at t_{s_1}. Then $p_{32}(t_{s_1}) > p_{34}(t_{s_1})$ and $p_{21}(t_{s_1}) > p_{24}(t_{s_1})$.

The second condition above ensures that the regime 27 is not preceded by regimes 20 and 19. The relationship $p_{14}(t) = \min\{p_{12}(t), p_{13}(t), p_{14}(t)\}$ can be verified to hold true $\forall t < t_{s_1}$. Since the function $p_{31}(t)$ is less than $p_{34}(t)$, $\forall t < t_{s_1}$, the regime 25 can not be preceded by regimes 21 and 24 in which $\alpha_{34}(t)$ is unity. It can be concluded that the regime 25 can not be preceded by regime 23 from the following arguments: Assume that regime 25 is preceded by regime 23 and let the instant of transition between regimes 25 and 23 be at $t = t_{s_2}$. Then $p_{31}(t_{s_2}) = p_{32}(t_{s_2})$ and $p_{23}(t_{s_2}) = p_{24}(t_{s_2})$. In regime 23 the dynamics of $p_{23}(t)$ and $p_{32}(t)$ are given by :

$$\begin{aligned}
\dot{p}_{23} &= -1 + a_{23}(p_{23} - p_{32}) \\
\dot{p}_{32} &= -1 + a_{32}(p_{32} - p_{23}) \\
\text{Therefore} \quad (\dot{p}_{23} - \dot{p}_{32}) &= (a_{23} + a_{32})(p_{23} - p_{32}) \\
\text{Hence} \quad p_{23}(t) - p_{32}(t) &= (p_{23}(t_{s_2}) - p_{32}(t_{s_2}))e^{(a_{23}+a_{32})(t-t_{s_2})}
\end{aligned}$$

If $p_{23}(t_{s_2}) - p_{32}(t_{s_2}) = 0$, then $(p_{23}(t) - p_{32}(t)) = 0$, $\forall t < t_{s_2}$ and consequently both the functions $p_{23}(t)$ and $p_{32}(t)$ have a slope equal to -1 in regime 23. If such is the case then

$$p_{23}(t) = p_{32}(t) \;=\; p_{23}(t_{s_2}) - (t - t_{s_2}) = p_{24}(t_{s_2}) - (t - t_{s_2})$$

But $p_{23}(t)$ as given by the above expression is greater than $p_{24}(t)$ for all $t < t_{s_2}$. Hence the routing variable $\alpha_{23}(t)$ can not be unity, which contradicts the assumption that the network operates in regime 23.
If $p_{23}(t_{s_2}) < p_{32}(t_{s_2})$ then it can be concluded that the slope of $p_{23}(t)$ is less than -1 and consequently

$$p_{23}(t) \;>\; p_{24}(t_{s_2}) - (t - t_{s_2}) > p_{24}(t)$$

and this leads to the same contradiction as above.
If $p_{23}(t_{s_2}) > p_{32}(t_{s_2})$ then the slope of $p_{32}(t)$ is less than -1 and therefore we get

$$\begin{aligned} p_{32}(t) &> p_{31}(t_{s_2}) - (t - t_{s_2}) \\ &> p_{31}(t) \quad \forall t < t_{s_2} \end{aligned}$$

Corresponding to this situation the routing variable $\alpha_{32}(t)$ can not be unity, which again contradicts the assumption that the network operates in regime 23.
Therefore regime 25 can not be preceded by regime 23. Thus the only regimes that could possibly precede regime 25 are regimes 19, 20, 22 and 26. The necessary and sufficient conditions for the regime 25 to be preceded by regime 19 can be stated as follows :

(i) $p_{31}(0) < p_{34}(0)$

(ii) Let the switching instant between $p_{31}(t)$ and $p_{34}(t)$ be at $t = t_{s_2}$. Then $p_{32}(t_{s_2})$ has to be greater than $p_{34}(t_{s_2})$.

Condition (ii) ensures that the regime 25 is not preceded by regime 20 or regime 26. The instant of switching between the regime 19 and 25 is obtained by solving the equation $p_{31}(t) = p_{34}(t)$ where $p_{31}(t)$ and $p_{34}(t)$ are given by the Equations (5.56) and (5.58) respectively. The regime 19 in turn may be preceded by other regimes. Let us investigate the various possibilities.
The arguments which were used to prove that the regime 23 can not be preceding regime 25 can be extended to show that regime 19 also can not be preceded by

regime 23. Since $p_{34}(t) > p_{31}(t)$ for all $t < t_{s_2}$, regime 19 can not be preceded by regimes 21 and 24 in which $\alpha_{34}(t)$ is unity. Similarly regimes 25 and 26 (in which $\alpha_{24}(t)$ is unity) can not be preceding regime 19. Thus the only two regimes that could possibly precede regime 19 are 20 and 22. In regime 19, the equations for $p_{32}(t)$, and $p_{23}(t)$ are given as the following:

$$\begin{aligned} p_{32}(t) &= p_{32}(t_{s_2})e^{a_{32}(t-t_{s_2})} + (1-e^{a_{32}(t-t_{s_2})})(1/a_{32}+1/a_{21}+1/a_{14}) \\ &+\frac{a_{14}a_{32}e^{a_{21}(t_{s_2}-T)}}{a_{21}(a_{14}-a_{21})(a_{21}-a_{32})}(e^{a_{21}(t-t_{s_2})}-e^{a_{32}(t-t_{s_2})}) \\ &+\frac{a_{21}a_{32}e^{a_{14}(t_{s_2}-T)}(e^{a_{14}(t-t_{s_2})}-e^{a_{32}(t-t_{s_2})})}{a_{14}(a_{21}-a_{14})(a_{14}-a_{32})} \end{aligned}$$

where $p_{32}(t_{s_2})$ is given as

$$p_{32}(t_{s_2}) = \frac{1-e^{a_{32}(t_{s_2}-T)}}{a_{32}}+\frac{1-e^{a_{24}(t_{s_2}-T)}}{a_{24}}+\frac{e^{a_{32}(t_{s_2}-T)}-e^{a_{24}(t_{s_2}-T)}}{a_{32}-a_{24}}.$$

$$\begin{aligned} p_{23}(t) &= p_{23}(t_{s_1})e^{a_{23}(t-t_{s_1})} + (1-e^{a_{23}(t-t_{s_1})})(1/a_{23}+1/a_{31}+1/a_{14}) \\ &+\frac{a_{14}a_{23}e^{a_{31}(t_{s_1}-T)}}{a_{31}(a_{14}-a_{31})(a_{31}-a_{23})}(e^{a_{31}(t-t_{s_1})}-e^{a_{23}(t-t_{s_1})}) \\ &+\frac{a_{31}a_{23}e^{a_{14}(t_{s_1}-T)}(e^{a_{14}(t-t_{s_1})}-e^{a_{23}(t-t_{s_1})})}{a_{14}(a_{31}-a_{14})(a_{14}-a_{23})} \end{aligned} \tag{5.76}$$

where $p_{23}(t_{s_1})$ is given as

$$p_{23}(t_{s_1}) = \frac{1-e^{a_{23}(t_{s_1}-T)}}{a_{23}}+\frac{1-e^{a_{34}(t_{s_1}-T)}}{a_{34}}+\frac{e^{a_{23}(t_{s_1}-T)}-e^{a_{34}(t_{s_1}-T)}}{a_{23}-a_{34}}$$

Consider the following four cases :

Case 1 : $p_{23}(0) < p_{21}(0)$ and $p_{32}(0) > p_{31}(0)$.

It is easy to infer that in this situation the regime 19 is preceded by regime 22 (which is the starting regime) and the switching instant between regime 22 and 19 is obtained by solving the equation $p_{23}(t) = p_{21}(t)$.

Case 2 : $p_{23}(0) > p_{21}(0)$ and $p_{32}(0) < p_{31}(0)$.

The regime 19 is preceded by regime 20 (starting regime) and the switching instant between regimes 20 and 19 is given by solving the equation $p_{32}(t) = p_{31}(t)$.

Case 3 : $p_{23}(0) > p_{21}(0)$ and $p_{32}(0) > p_{31}(0)$.

The regime 19 itself is the starting regime.

Case 4 : $p_{23}(0) < p_{21}(0)$ and $p_{32}(0) < p_{31}(0)$.

Let the switching instant between $p_{21}(t)$ and $p_{23}(t)$ be denoted by t_{13} and let that between $p_{32}(t)$ and $p_{31}(t)$ be denoted by t_{12}. If $t_{13} > t_{12}$, then the regime 19 is preceded by regime 22 (which is the starting regime), while if $t_{12} > t_{13}$ then the regime 19 is preceded by regime 20 (starting regime).

As mentioned above, the regime 25 could possibly preceded by regime 20. The necessary and sufficient condition under which regime 25 is preceded by regime 20 can be stated as follows :

(i) $p_{32}(0) > p_{31}(0)$ where $p_{32}(t)$ and $p_{31}(t)$ are as given by Equations (5.57) and (5.56) respectively.

(ii) $p_{21}(0) < p_{24}(0)$ where $p_{21}(t)$ is given by the Equation (5.53).

(iii) Let the instant of switching of $p_{32}(t)$ and $p_{31}(t)$ be at $t = t_{s_2}$. Then $p_{21}(t_{s_2}) = p_{24}(t_{s_2})$.

Such a regime 20 which precedes regime 25 has to be the starting regime.

The other two regimes that could possibly precede regime 25 are regimes 22 and 26. The necessary and sufficient conditions for regime 25 to be preceded by regime 22 are the following :

(i) $p_{23}(0) < p_{24}(0)$ where $p_{23}(t)$ and $p_{24}(t)$ are given by the Equations (5.76) and (5.55) respectively.

(ii) Let the switching instant between $p_{23}(t)$ (as given by the Equation (5.76)) and $p_{24}(t)$ (as given by the Equation (5.55)) be denoted by t_{s_2}. Then it is necessary that $p_{21}(t_{s_2})$ is greater than $p_{24}(t_{s_2})$.

(iii) At $t = t_{s_2}$, $p_{32}(t)$ (as given by the Equation (5.57)) is greater than $p_{31}(t)$ (as given by the Equation (5.56)).

The condition (ii) above ensures that regime 25 is not preceded by regime 19 and 20 and condition (iii) ensures that regime 25 is not preceded by regime 26.

The regime 22 may in turn be preceded by regime 19 (It can be argued that the other regimes can not precede the regime 22). The necessary and sufficient condition for the regime 22 to be preceded by regime 19 is that $p_{21}(0) < p_{23}(0)$ where $p_{23}(t)$ is given by the Equation (5.76) and $p_{21}(t)$ is given by the Equation (5.53). The switching instant t_{s_3} between the regimes 22 and 19 is obtained by solving the equation $p_{23}(t) = p_{21}(t)$.

The only regime which could possibly precede the regime 19 above is regime 20. The necessary and sufficient condition for this to take place is that $p_{32}(0) < p_{31}(0)$ where $p_{31}(t)$ is given by the Equation (5.56) and $p_{32}(t)$ is given below.

$$\begin{aligned} p_{32}(t) &= p_{32}(t_{s_3})e^{a_{32}(t-t_{s_3})} + (1-e^{a_{32}(t-t_{s_3})})(1/a_{32}+1/a_{21}+1/a_{14}) \\ &+ \frac{a_{14}a_{32}e^{a_{21}(t_{s_3}-T)}}{a_{21}(a_{14}-a_{21})(a_{21}-a_{32})}(e^{a_{21}(t-t_{s_3})} - e^{a_{32}(t-t_{s_3})}) \\ &+ \frac{a_{21}a_{32}e^{a_{14}(t_{s_3}-T)}(e^{a_{14}(t-t_{s_3})} - e^{a_{32}(t-t_{s_3})})}{a_{14}(a_{21}-a_{14})(a_{14}-a_{32})}. \end{aligned} \tag{5.77}$$

To obtain $p_{32}(t_{s_3})$ in the above equation, note that during the interval $[t_{s_3}, t_{s_2}]$, the dynamics of $p_{32}(t)$ is given as

$$\dot{p}_{32}(t) = -1 + a_{32}(p_{32} - p_{23})$$

where $p_{23}(t)$ is given by the Equation (5.76). Therefore $p_{32}(t_{s_3})$ is given as

$$\begin{aligned} p_{32}(t_{s_3}) &= p_{32}(t_{s_2})e^{a_{32}(t_{s_3}-t_{s_2})} + (1/a_{32}+1/a_{23}+1/a_{31})(1-e^{a_{32}(t_{s_3}-t_{s_2})}) \\ &+ \frac{(e^{a_{23}(t_{s_3}-t_{s_2})} - e^{a_{32}(t_{s_3}-t_{s_2})})}{(a_{23}-a_{32})}e^{a_{23}(t_{s_2}-t_{s_1})}\left(\frac{a_{14}a_{23}e^{a_{31}(t_{s_1}-T)}}{a_{31}(a_{14}-a_{31})(a_{23}-a_{31})}\right. \\ &\left. + p_{23}(t_{s_1}) - (1/a_{23}+1/a_{31}+1/a_{14}) - \frac{a_{31}a_{23}e^{a_{14}(t_{s_1}-T)}}{a_{14}(a_{31}-a_{14})(a_{14}-a_{23})}\right) \\ &- \frac{a_{32}e^{a_{31}(t_{s_2}-t_{s_1})}a_{14}a_{23}e^{a_{31}(t_{s_1}-T)}}{(a_{31}-a_{32})a_{31}(a_{14}-a_{31})(a_{31}-a_{23})}(e^{a_{31}(t_{s_3}-t_{s_2})} - e^{a_{32}(t_{s_3}-t_{s_2})}) \\ &- \frac{a_{32}e^{a_{14}(t_{s_2}-t_{s_1})}a_{31}a_{23}e^{a_{14}(t_{s_1}-T)}}{(a_{14}-a_{32})a_{14}(a_{31}-a_{14})(a_{14}-a_{23})}(e^{a_{14}(t_{s_3}-t_{s_2})} - e^{a_{32}(t_{s_3}-t_{s_2})}). \end{aligned}$$

In the above equation $p_{23}(t_{s_1})$ and $p_{32}(t_{s_2})$ are obtained by evaluating the Equations (5.54) and (5.57) at the corresponding instants.

The switching instant t_{s_4} between the regime 20 and regime 19 is obtained by solving the equation $p_{32}(t) = p_{31}(t)$ where $p_{32}(t)$ is given by Equation (5.77) and $p_{31}(t)$ is given by Equation (5.56).

The above discussion ends the case where in regime 25 is preceded by regime 22 and which in turn is preceded by other regimes. Let us now investigate the case wherein the terminal regime 25 is preceded by regime 26.

The necessary and sufficient conditions under which the regime 25 is preceded by regime 26 can be stated as follows :

(i) $p_{32}(0) < p_{31}(0)$ where $p_{32}(t)$ and $p_{31}(t)$ are given by the Equations (5.57) and (5.56) respectively.

(ii) Let t_{s_2} denote the instant at which the functions $p_{32}(t)$ and $p_{31}(t)$ as given by the Equations (5.57) and (5.56) intersect. Then

  (a) $p_{23}(t_{s_2}) > p_{24}(t_{s_2})(t)$

  (b) $p_{21}(t_{s_2}) > p_{24}(t_{s_2})(t)$

  where $p_{23}(t)$ and $p_{21}(t)$ are given by the Equations (5.76) and (5.53) respectively.

If both the conditions (i) and (ii) above are satisfied, then the regime 25 is preceded by regime 26. It can be argued that the relationships

(i) $p_{14}(t) = \min\{p_{14}(t), p_{12}(t), p_{13}(t)\}$

(ii) $p_{32}(t) = \min\{p_{31}(t), p_{32}(t), p_{34}(t)\}$

hold true $\forall t < t_{s_2}$. Therefore regime 26 can not be preceded by any regime in which the routing variable $\alpha_{14}(t)$ and/or $\alpha_{32}(t)$ is zero. Thus the only regime that could possibly precede the regime 26 is regime 20. It can be inferred that the necessary and sufficient condition for the regime 26 to be preceded by regime 20 is that $p_{21}(0) < p_{24}(0)$ and the switching instant t_{s_3} between the regimes 26 and 20 is obtained by solving the equation $p_{21}(t) = p_{24}(t)$. Here $p_{21}(t)$ and $p_{24}(t)$ are given by the Equations (5.53) and (5.55) respectively. It can also be easily argued that such a regime 20 has to be the starting regime.

The above discussion exhausts all the possibilities when the terminal regime 27 is preceded by regime 25. So far, we investigated the cases wherein regime 27 is preceded by regimes 19, 20, 21, 22 and 25. Let us now examine the case wherein the regime 27 is preceded by regime 26.

The necessary and sufficient conditions for the regime 27 to be preceded by regime 26 are the following :

(i) $p_{32}(0) < p_{34}(0)$ where $p_{32}(t)$ and $p_{34}(t)$ are given by the Equations (5.57) and (5.58) respectively.

(ii) Let the instant at which the functions $p_{32}(t)$ and $p_{34}(t)$ intersect be at t_{s_1}. Then

   (a) $p_{21}(0) > p_{24}(0)$

   (b) $p_{31}(0) > p_{34}(0)$

The regime 26 may in turn be preceded by other regimes. Based on the properties of the costate variables (which we have so far exploited), it can be argued that the only regimes that could possibly precede the regime 26 are regimes 19, 20 and 25. The necessary and sufficient conditions for the regime 26 to be preceded by regime 19 are the following :

(i) $p_{31}(0) < p_{32}(0)$ where $p_{31}(t)$ and $p_{32}(t)$ are given by the Equations (5.56) and (5.57) respectively.

(ii) $p_{21}(0) < p_{24}(0)$ where $p_{21}(t)$ and $p_{24}(t)$ are given by Equations (5.53) and (5.55) respectively.

(iii) Let the instant at which the functions $p_{31}(t)$ and $p_{32}(t)$ intersect be denoted by t_{s_2}. It is necessary that $p_{21}(t)$ and $p_{24}(t)$ intersect at the same instant. i.e. $p_{21}(t_{s_2}) = p_{24}(t_{s_2})$.

If all the three conditions stated above are satisfied, then the regime 26 is preceded by regime 19. The instant of transition between regimes 19 and 26 is obtained by solving the equation $p_{21}(t) = p_{24}(t)$. Such a regime 19 can in turn be preceded by

other regimes. Based on the properties of the costate variables in such a regime 19 (that precedes the regime 26), it can be argued out that the only regimes that could possibly precede the regime 19 are 20 and 22. The necessary and sufficient condition for the regime 19 to be preceded by regime 20 is given as

- $p_{32}(0) < p_{31}(0)$ where the expression for $p_{32}(0)$ given as

$$\begin{aligned} p_{32}(0) &= p_{32}(t_{s_2})e^{-a_{32}t_{s_2}} + (1 - e^{-a_{32}t_{s_2}})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ &+ \frac{a_{14}a_{32}e^{a_{21}(t_{s_2}-T)}}{a_{21}(a_{14}-a_{21})(a_{21}-a_{32})}(e^{-a_{21}t_{s_2}} - e^{-a_{32}t_{s_2}}) \\ &+ \frac{a_{21}a_{32}e^{a_{14}(t_{s_2}-T)}(e^{-a_{14}t_{s_2}} - e^{-a_{32}t_{s_2}})}{a_{14}(a_{21}-a_{14})(a_{14}-a_{32})} \end{aligned} \tag{5.78}$$

and $p_{32}(t_{s_2})$ in the above Equation (5.78) is given as

$$p_{32}(t_{s_2}) = \frac{1 - e^{a_{32}(t_{s_2}-T)}}{a_{32}} + \frac{1 - e^{a_{24}(t_{s_2}-T)}}{a_{24}} + \frac{e^{a_{32}(t_{s_2}-T)} - e^{a_{24}(t_{s_2}-T)}}{a_{32}-a_{24}}.$$

The expression for $p_{31}(0)$ is given as

$$p_{31}(0) = \frac{1 - e^{-a_{31}T}}{a_{31}} + \frac{1 - e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{31}-T} - e^{-a_{14}T}}{a_{31}-a_{14}}.$$

The necessary and sufficient condition for the regime 19 (that precedes regime 26) to be preceded by regime 22 is given as

- $p_{23}(0) < p_{21}(0)$ where $p_{23}(0)$ is given as

$$\begin{aligned} p_{23}(0) &= p_{23}(t_{s_2})e^{-a_{23}t_{s_2}} + (1 - e^{-a_{23}t_{s_2}})(1/a_{23} + 1/a_{31} + 1/a_{14}) \\ &+ \frac{a_{14}a_{23}e^{a_{31}(t_{s_2}-T)}}{a_{31}(a_{14}-a_{31})(a_{31}-a_{23})}(e^{-a_{31}t_{s_2}} - e^{-a_{23}t_{s_2}}) \\ &+ \frac{a_{31}a_{23}e^{a_{14}(t_{s_2}-T)}(e^{-a_{14}t_{s_2}}) - e^{-a_{23}t_{s_2}})}{a_{14}(a_{31}-a_{14})(a_{14}-a_{23})} \end{aligned} \tag{5.79}$$

where $p_{23}(t_{s_2})$ in the Equation (5.79) above is given as

$$\begin{aligned} p_{23}(t_{s_2}) &= p_{23}(t_{s_1})e^{a_{23}(t_{s_2}-t_{s_1})} + (1 - e^{a_{23}(t_{s_2}-t_{s_1})})(1/a_{23} + 1/a_{32} + 1/a_{24}) \\ &+ \frac{a_{24}a_{23}e^{a_{32}(t_{s_1}-T)}}{a_{32}(a_{24}-a_{32})(a_{32}-a_{23})}(e^{a_{32}(t_{s_2}-t_{s_1})} - e^{a_{23}(t_{s_2}-t_{s_1})}) \\ &+ \frac{a_{32}a_{23}e^{a_{24}(t_{s_1}-T)}(e^{a_{24}(t_{s_2}-t_{s_1})} - e^{a_{23}(t_{s_2}-t_{s_1})})}{a_{24}(a_{32}-a_{24})(a_{24}-a_{23})}. \end{aligned} \tag{5.80}$$

Note that $p_{23}(t_{s_1})$ in the Equation (5.80) above is given by

$$p_{23}(t_{s_1}) = \frac{1 - e^{a_{23}(t_{s_1}-T)}}{a_{23}} + \frac{1 - e^{a_{34}(t_{s_1}-T)}}{a_{34}} + \frac{e^{a_{23}(t_{s_1}-T)} - e^{a_{34}(t_{s_1}-T)}}{(a_{23} - a_{34})}.$$

And the term $p_{21}(0)$ is given by

$$p_{21}(0) = \frac{1 - e^{-a_{21}T}}{a_{21}} + \frac{1 - e^{-a_{14}T}}{a_{14}} + \frac{e^{-a_{21}T} - e^{-a_{14}T}}{(a_{21} - a_{14})}$$

It can be argued that regimes 20 and 22 which precede the regime 19 are the starting regimes.

The necessary and sufficient conditions under which the regime 26 is preceded by regime 20 can be stated as follows:

(i) $p_{21}(0) < p_{24}(0)$ where $p_{21}(t)$ and $p_{24}(t)$ are given by the Equations (5.53) and (5.55) respectively.

(ii) Let the instant at which $p_{21}(t)$ and $p_{24}(t)$ intersect be at $t = t_{s_2}$. Then $p_{32}(t_{s_2}) < p_{31}(t_{s_2})$ where $p_{32}(t)$ and $p_{31}(t)$ are given by the Equations (5.57) and (5.56) respectively.

Assume that the above conditions are satisfied and therefore the regime 26 is preceded by regime 20. Based on the properties of the costate variables in the regime 20, it can be argued that the only regime that could possibly precede the regime 20 is regimes 19. The necessary and sufficient conditions under which regime 20 is preceded by regime 19 is given as :

- $p_{31}(0) < p_{32}(0)$ where $p_{31}(t)$ is given by the Equation (5.56) and $p_{32}(t)$ is given by the following expression :

$$\begin{aligned} p_{32}(t) &= p_{32}(t_{s_2})e^{a_{32}(t-t_{s_2})} + (1 - e^{a_{32}(t-t_{s_2})})(1/a_{32} + 1/a_{21} + 1/a_{14}) \\ &+ \frac{a_{14}a_{32}e^{a_{21}(t_{s_2}-T)}}{a_{21}(a_{14} - a_{21})(a_{21} - a_{32})}(e^{a_{21}(t-t_{s_2})} - e^{a_{32}(t-t_{s_2})}) \\ &+ \frac{a_{21}a_{32}e^{a_{14}(t_{s_2}-T)}(e^{a_{14}(t-t_{s_2})} - e^{a_{32}(t-t_{s_2})})}{a_{14}(a_{21} - a_{14})(a_{14} - a_{32})}. \end{aligned}$$

where $p_{32}(t_{s_2})$ above is obtained by evaluation the expression (5.57) at $t = t_{s_2}$.

The regime 19 will in turn be preceded by regime 22 if $p_{23}(0) < p_{21}(0)$ where $p_{23}(t)$ is given by the Equation (5.76) and $p_{21}(t)$ is given by the Equation (5.53). The instant of switching from regime 22 to regime 19 is obtained by solving the equation $p_{23}(t) = p_{21}(t)$. It can be argued that the regime 22 is the starting regime.
Finally let us consider the case wherein the regime 26 is preceded by regime 25. The necessary and sufficient conditions for the regime 26 to be preceded by regime 25 can be stated as follows :

(i) $p_{31}(0) < p_{32}(0)$ where $p_{31}(t)$ and $p_{32}(t)$ are given by the Equations (5.56) and (5.57) respectively.

(ii) Let the instant of switching between $p_{31}(t)$ and $p_{32}(t)$ be at $t = t_{s_2}$. Then $p_{21}(t_{s_2}) > p_{24}(t_{s_2})$.

The regime 25 in turn may be preceded by other regimes. Based on the properties of the costate variables in the regime 25, it can be argued that the only regimes that could possibly precede the regime 25 are 19 and 22. The necessary and sufficient conditions for the regime 25 to be preceded by regime 19 are the following :

(a) $p_{21}(0) < p_{24}(0)$ where $p_{21}(t)$ and $p_{24}(t)$ are given by the Equations (5.53) and (5.55) respectively .

(b) Let the instant at which the functions $p_{21}(t)$ and $p_{24}(t)$ intersect be denoted by t_{s_3}. Then $p_{23}(t_{s_3})$ has to be greater $p_{24}(t_{s_3})$.

The condition (b) above ensures that the regime 25 is not preceded by regime 22. If (b) is violated then the regime 25 is preceded by regime 22. It can be argued that the only regime that possibly precedes regime 19 is regime 22. The (necessary and sufficient) condition under which the regime 19 is further preceded by regime 22 is $p_{23}(0) < p_{21}(0)$ where $p_{23}(t)$ is given by the Equation (5.76). Such a regime 22 can be argued to be the starting regime.
The necessary and sufficient conditions under which the regime 25 is preceded by regime 22 are given as the following :

(i) $p_{23}(0) < p_{24}(0)$ where $p_{23}(t)$ is given by the Equation (5.76).

(ii) Let the instant at which $p_{23}(t)$ and $p_{24}(t)$ intersect be denoted by t_{s_3}. Then $p_{21}(t_{s_3})$ is greater than $p_{24}(t_{s_3})$.

It can be argued, based on the properties of the costate variables in regime 22 (that precedes regime 25), that the only regime that could possibly precede the regime 22 is regime 19. The (necessary and sufficient) condition under which this takes place is given by $p_{21}(0) < p_{23}(0)$ where $p_{23}(t)$ is given by the Equation (5.76).
The above discussion exhausts all the possible regime transitions and the switching instants for the various cases when the link parameters for the direct links (1,4), (2,4) and (3,4) satisfy the relationship $a_{14} > a_{24} > a_{34}$. Based on the above discussion the possible regime transitions for this case is shown in the Figure 5.30.

The following observations can be made regarding the optimal routing strategy for this case (wherein $a_{14} > a_{24} > a_{34}$) :

(i) Traffic arriving at node 1 is routed onto the direct link (1,4) (which is the *fastest* direct link) for the entire duration of network operation.

(ii) The network operation never enters regime 23 in which the both the routing variables α_{23} and α_{32} are unity (**loop-free** property).

(iii) The network operation never enters regime 24 in which the traffic arriving at the node 2 is sent on the path $\{(2,3),(3,4)\}$ to the destination. This can be justified since the link (3,4) is *slower* than link (2,4) and intuitively it appears unreasonable to use an alternate path (at node 2) consisting of a slower link than the direct link to the destination.

The observations (ii) and (iii) above can be extended to the case where $a_{14} > a_{34} > a_{24}$ and it can be concluded that the network operation never enters regimes 23 and 26 in this case.

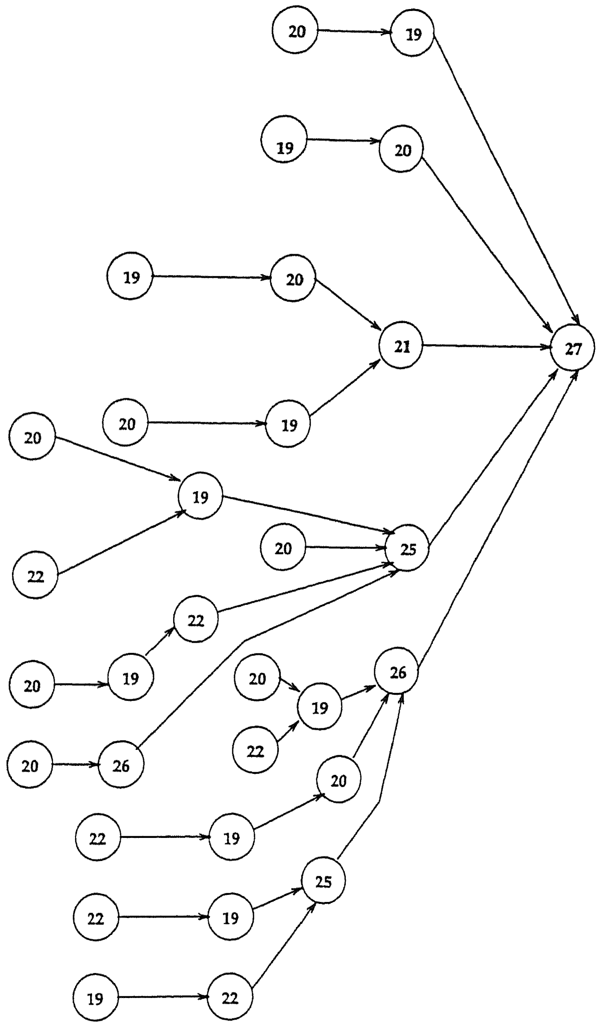

Figure 5.30: The possible regime transitions for the case $a_{14} > a_{24} > a_{34}$

Remark 5.5.1 : From reasonings similar to the ones employed in the above discussion, the following observations can be made regarding the case wherein $a_{24} > a_{14} > a_{34}$:

(i) The traffic arriving at node 2 is routed onto the direct link (2,4) for the entire duration of network operation i.e. $\alpha_{24}(t)$ equals unity for all $t \in [0, T]$. Consequently the regimes $1, 2, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 19, 20, 21, 22, 23$ and 24 are ruled out.

(ii) The network operation never enters regimes 16 and 18.

Similarly for the case wherein $a_{24} > a_{34} > a_{14}$, the network operation never enters regimes 1, 2, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 24, and 25.
For the case wherein $a_{34} > a_{24} > a_{14}$, the network operation never enters regimes 1, 2, 3, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 21, 22, 23, 25, and 26.
For the case wherein $a_{34} > a_{14} > a_{24}$, the network operation never enters regimes 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25 and 26.

5.6 Suboptimal Strategy

Our analysis in the previous sections shows that the *fastest* direct link is used for the entire duration of the network operation. Furthermore, under certain additional constraints on the link parameters of the network which we stated earlier, the optimal routing strategy for the composite network was shown to be the same as the one synthesised from the *locally* optimal routing strategies for the individual units. When these constraints are violated, the possible regime transitions and the equations to specify them were arrived at. A possible way of suboptimally routing the packets, which incorporates features of the optimal strategy, and therefore can intuitively be expected to achieve a good performance is as follows :

- At the source node which has the *fastest* direct link, route all the incoming the traffic onto this direct link for the entire duration of network operation.

- Consider the two network (three node) units which share the *fastest* direct link. Synthesise the *locally* optimal routing strategy for these units.

In Table 5.8, we give the performance of this suboptimal strategy with the optimal strategy for some illustrative examples. J_1 is the performance of the optimal strategy and J_2 that of the suboptimal strategy. We consider a traffic pattern $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = 10, \forall t \in [0, 10]$. The network operation is assumed to be for a duration of 10 units and the initial buffer occupancies are assumed to be zero. The optimal strategy is obtained by numerically solving the differential equations (5.41)-(5.49) in the costate variables.

Table 5.8: Optimal and Suboptimal Strategies for the Four Node Network: Numerical Examples

Link parameters	Optimal Strategy	Suboptimal Strategy	J_1	J_2
$a_{12} = a_{32} = 0.8, a_{13} = 0.9$ $a_{23} = a_{21} = a_{31} = 0.6,$ $a_{14} = 0.7, a_{24} = a_{34} = 0.3$	$\alpha_{21} = \alpha_{31} = 1$ in $[0, 2.0],$ $\alpha_{24} = \alpha_{34} = 1$ in $(2, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	$\alpha_{21} = \alpha_{31} = 1$ in $[0, 2.0],$ $\alpha_{24} = \alpha_{34} = 1$ in $(2, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	615.135	615.135
$a_{12} = a_{32} = 0.6$ $a_{34} = a_{14} = 0.25$ $a_{24} = 0.8,\ a_{21} = 0.5$ $a_{31} = 0.3, a_{23} = 0.1, a_{13} = 0.7$	$\alpha_{12} = \alpha_{32} = 1$ in $[0, 6.5]$ $\alpha_{14} = \alpha_{34} = 1$ in $[6.5, 10]$ $\alpha_{24} = 1$ in $[0, 10]$	$\alpha_{12} = \alpha_{32} = 1$ in $[0, 6.5]$ $\alpha_{14} = \alpha_{34} = 1$ in $[6.5, 10]$ $\alpha_{24} = 1$ in $[0, 10]$	498.887	498.887
$a_{12} = a_{13} = 0.1$ $a_{23} = a_{32} = 0.2$ $a_{24} = a_{34} = 0.3,\ a_{21} = 0.7$ $a_{31} = 0.6, a_{14} = 0.8$	$\alpha_{21} = \alpha_{31} = 1$ in $[0, 4.2]$ $\alpha_{21} = \alpha_{34} = 1$ in $(4.2, 5.9]$ $\alpha_{24} = \alpha_{34} = 1$ in $(5.9, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	$\alpha_{21} = \alpha_{31} = 1$ in $[0, 4.2]$ $\alpha_{21} = \alpha_{34} = 1$ in $(4.2, 5.9]$ $\alpha_{24} = \alpha_{34} = 1$ in $(5.9, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	562.05244	562.05244
$a_{14} = 0.6, a_{24} = a_{34} = 0.25,$ $a_{21} = 0.85, a_{31} = 0.5, a_{32} = 0.9$ $a_{12} = a_{23} = a_{13} = 0.1$	$\alpha_{32} = \alpha_{21} = 1$ in $[0, 2.1]$ $\alpha_{31} = \alpha_{21} = 1$ in $(2.1, 4.32]$ $\alpha_{21} = \alpha_{34} = 1$ in $(4.32, 6.84]$ $\alpha_{34} = \alpha_{24} = 1$ in $(6.84, 10]$ $\alpha_{14} = 1,$ in $[0, 10]$	$\alpha_{31} = \alpha_{21} = 1$ in $[0, 1.2]$ $\alpha_{34} = \alpha_{21} = 1$ in $(1.2, 6.84]$ $\alpha_{34} = \alpha_{24} = 1$ in $(6.84, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	558.321	563.66382
$a_{14} = 0.6, a_{24} = a_{34} = 0.25$ $a_{31} = 0.8, a_{21} = 0.55, a_{23} = 0.9$ $a_{12} = a_{32} = a_{13} = 0.1$	$\alpha_{23} = \alpha_{31} = 1$ in $[0, 1.73]$ $\alpha_{21} = \alpha_{31} = 1$ in $(1.73, 4.02]$ $\alpha_{24} = \alpha_{31} = 1$ in $(4.02, 6.54]$ $\alpha_{24} = \alpha_{34} = 1$ in $(6.54, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	$\alpha_{21} = \alpha_{31} = 1$ in $[0, 3.12]$ $\alpha_{24} = \alpha_{31} = 1,$ in $(3.12, 6.54]$ $\alpha_{24} = \alpha_{34} = 1,$ in $(6.54, 10]$ $\alpha_{14} = 1$ in $[0, 10]$	589.4321	595.18413

5.7 Conclusions

In this chapter, we investigated the problem of optimal routing in some networks composed of the primitive units of the Chapter 3 and Chapter 4. We first considered a topology consisting of the two node network units where in m units (of the total n units which constitute the network) have a link (the *faster* link of the two) of finite channel capacity and the remaining $(n - m)$ units have both the links of infinite channel capacity. We showed that in the units which have both the links of infinite channel capacity, all the incoming traffic is routed onto the *faster* link for the entire duration of network operation. In the other network units, there could be intervals of *partial routing* during which, a fraction of the incoming traffic equal to the ratio of the channel capacity of the link to the total incoming traffic to the network unit, is sent on the *faster* link. The properties of the optimal routing strategy when these units are considered in isolation were shown to hold true even when they are considered in conjunction with the other units as in the topology considered. For the case in which only one unit (i.e. $m = 1$) has a link of finite channel capacity, we obtained the set of equations required to specify the optimal routing strategy. Solution to these equations requires the knowledge of the load pattern for the entire duration of the network operation and hence necessitates an off-line computation. We therefore proposed an on-line implementable suboptimal strategy for this network. In Section 5.2.3 we compared the performance of this suboptimal strategy with that of the optimal strategy and the performance (which is the best possible achievable) of the routing strategy if all the links of the network were of infinite channel capacity, for the case where $m = 1, n = 3$.

We then considered a network topology consisting of the three node primitives of the Chapter 4 in the Section 5.3. When all the links of this topology are of infinite channel capacity, then the optimal routing strategy for this network is *bang-bang* in nature and is independent of the input traffic. Since this network topology is viewed as being composed of simpler three node units, and since the optimal ategy for the individual units (*locally* optimal to the units) can be specified

in terms of a single switching instant t_s (as shown in the Chapter 4), it is meaningful to compare the performance of such a suboptimal strategy (synthesised from the *locally* optimal strategies) with respect to the performance of the *globally* optimal strategy. We compared the performance of these two strategies for some typical values of the link parameters of the network. When one of the links of a unit has a direct link of finite channel capacity, a scheme for ensuring that the network operates in the linear mode is to reroute the excess traffic (which drives the link of finite channel capacity) to the next unit in the same layer of the topology. We compared the performance of the network (for various link parameters) operated under the *globally* optimal and the *locally* optimal strategies with this scheme of restricting to the linear mode, implemented.

Finally in Section 5.4 we investigated the problem of optimal routing in a four-node network when all its links are of infinite channel capacity. We proved that the network operation ends with a direct routing of the incoming traffic at all the three source node. At the source node which has the *fastest* direct link, the incoming traffic is routed onto this link for the entire duration of network operation. We showed that under certain conditions on the link parameters, the optimal routing strategy for this network is the same as the one synthesised from the *locally* optimal routing strategies for the constituent units. When these conditions are violated, a suboptimal strategy in terms of *locally* optimal routing strategy for the constituent units which share the *fastest* direct link was proposed and its performance was compared with the optimal strategy in the case of some illustrative examples.

Chapter 6

Conclusions

The problem of synthesising routing strategies which optimize the performance of a network has drawn much attention in the recent times. A large number of them, each one claiming to be better than the other, have been suggested by various authors in this context. Based on how these routing strategies incorporate the variations in the traffic and/or the network conditions, they can be broadly classified into *static*, *quasistatic* and *dynamic routing strategies*. In *static routing*, the fractions of the traffic routed onto the outgoing links at each node of the network are fixed in time, and are decided upon before the network starts operation. In *quasistatic routing*, changes in routes take place at given instants of time and/or whenever extreme conditions such as a link failure occur within the network. *Dynamic routing* incorporates continuous changing of routes depending upon the instantaneous system states and traffic conditions. The growing need to incorporate adaptivity to the variations in the traffic and/or the network conditions in order to improve the performance (of the network) has resulted in a shift in trend from static routing to dynamic routing strategies. Such a shift is perceived in the context of circuit-switched networks also.

A survey of the literature on routing suggests that intuition, heuristics and simulation rather than a thorough analysis, have often been the guiding factors in the choice of the routing algorithms in many of the existing networks. While

this may be attributed, in part, to the ever increasing need of network community to come up with practically more efficient routing strategies, it also motivates the need to have theoretical frameworks for the study of them. A major (theoretical) contribution in this direction in terms of providing a systematic approach to the synthesis of flow control and routing strategies was made by Filipiak in [21]. The approach adopted by Filipiak was to model the network as a dynamical system using flow models. The basic model of dynamic flows relates the growth in the amount of packets/messages in the system at time t, by means of a deterministic differential equation, to the intensity of newly arriving traffic, the intensity of traffic discarded/rejected by the system and the intensity of successfully delivered traffic. It was assumed that the intensity of the outgoing traffic can be approximated by a non-linear function of the system state. The investigation in this thesis started within this framework. We assumed that the flow out from any buffer onto the link, will increase with increasing buffer occupancy and will saturate at a value equal to the channel capacity of the link. Furthermore, when the buffer is empty the flow out will be zero. Based on these assumptions, we considered a model in which the flow out function increases linearly with increase in buffer occupancy till it reaches the value equal to the channel capacity of the link. Since practical networks are subjected to shut downs and need rebooting from time to time, we considered a finite time duration operation of the network.

From the consideration that optimal (or at least good suboptimal) routing strategies for large networks can be synthesised from those for simpler network units that constitute them, we investigated the problem of optimal routing in two such simple units, namely a two node unit in Chapter 3 and a three node unit in Chapter 4. We then considered network topologies which are composed of these units in Chapter 5. The major results which we have derived in this thesis regarding the nature of optimal routing in these networks can be stated as the following:

We first investigated the problem of optimal routing in a two node network which has a set of two parallel links between the source and the destination. The *faster* link of the two was assumed to be of finite channel capacity while the *slower*

We argued that the optimal routing strategy has the **loop-free** property in this case too. Of the 27 modes in which the network can operate the optimal routing strategy admits only four of these in a terminal interval.

We then considered the situation where in one of the direct links has finite channel capacity. However, the link parameters are such that, if all the links were of infinite channel capacity, then the optimal routing strategy would be a direct routing at both the source nodes for the entire duration of network operation. Under the additional assumption that the buffer occupancy for the link with finite channel capacity is below the saturation value, we derived a set of equations in terms of the link parameters and the input traffic to the network, required to specify the optimal routing strategy. Since the solution to this set of equations requires the knowledge of the load pattern for the entire duration of network operation and this, in turn, necessitates an off-line computation, an on-line implementable suboptimal strategy was proposed. The optimal and suboptimal strategies were illustrated with examples and their performances were also compared.

We also considered an exponential model for the flow out function for the three node topology. We conducted some preliminary numerical investigations from which certain interesting observations were made. The optimal routing strategy was found to be having the **loop-free** property. It was also observed that the network operation always ends with a direct routing at both the source nodes. The optimal routing strategy was found to be having the single switching property. However, we did not investigate whether these properties can be analytically proved and/or can be made use of in specifying the routing strategy by means of a simple procedure.

In Chapter 5, we investigated the problem of optimal routing in some networks which are composed of the network units of the Chapter 3 and Chapter 4. We first considered a topology consisting of the two node network units where in m units, of the total n which constitute the topology, have a link of finite channel capacity. The *faster* link of the two was assumed to be of finite channel capacity while the *slower* link was assumed to be of infinite channel capacity. We showed

that in the units which have both the links of infinite channel capacity, all the incoming traffic is routed onto the *faster* link for the entire duration of network operation. In the other network units, there could be intervals of *partial routing* during which, a fraction of the incoming traffic equal to the ratio of the channel capacity of the link to the total incoming traffic to the network unit, is sent on the *faster* link. The properties of the optimal routing strategy when these network units were considered in isolation were shown to hold true even when they are considered in conjunction with the other units as in the topology considered. For the case in which only one unit has a link of finite channel capacity, we derived a set of equations (in terms of the link parameters and the input traffic) required to specify the optimal routing strategy. Solution to these equations requires the knowledge of the load pattern for the entire duration of network operation and this, in turn, necessitates an off-line computation. We therefore proposed an on-line implementable suboptimal strategy. The optimal and suboptimal strategies were illustrated by means of examples and their performances were also compared.

We considered a network topology consisting of the three node primitives of Chapter 4 in Section 5.3. When all the links of this unit are of infinite channel capacities, the optimal routing strategy is *bang-bang* in nature and is independent of the input traffic. Since this network topology is viewed as being composed of simpler three node units, and since the optimal routing strategy for the individual units (*locally optimal* to the units) can be specified in terms of a single switching instant t_s, we compared the performance of the *globally optimal* strategy with that of such a suboptimal strategy (which is synthesised from *locally optimal* strategies for the constituent units), in the case of some illustrative examples. We also considered a scheme of restricting the traffic so that the network operates in the linear mode, even when one of the direct links of one of the network unit has a finite channel capacity.

Finally, we investigated the problem of optimal routing in a four-node hub network when all its links are of infinite channel capacity. We proved that the network operation ends with a direct routing of the incoming traffic at all the three

source node. At the source node which has the *fastest* direct link, the incoming traffic is routed onto this link for the entire duration of network operation. We also showed that under certain conditions on the link parameters, the optimal routing strategy for this network is the same as the one synthesised from the *locally optimal* routing strategies for the constituent units. When these conditions are violated, a suboptimal strategy in terms of a *locally optimal* routing strategy for the constituent units (which share the *fastest* direct link was proposed), and its performance was compared with the optimal strategy in the case of some illustrative examples.

From our studies in this thesis, several interesting possibilities for future work emerge. Some of these can be stated as the following:

- We have not investigated whether there are any computationally efficient procedures to solve the system of equations which we derived (for the two node unit in Chapter 3, for the three node unit in Chapter 4 and for the composite network of two node units in Chapter 5) for specifying the optimal routing strategy. An investigation along this direction would make this study complete.

- We have suggested a suboptimal algorithm for the two node network in Chapter 3, for the three node network unit (under certain restrictions on the link parameters) in Chapter 4 and for the composite network consisting of the two node units in Chapter 5. No quantitative statements about the performance of these suboptimal algorithms have been made. It would be interesting to investigate whether any bounds on the performance can be obtained analytically.

- Our analysis for the three node unit ends with the case of a single link with finite channel capacity. When more than one link has finite channel capacity, it would be interesting to see if optimal routing strategy has the loop-free property. It would also be worthwhile investigating whether a good suboptimal algorithm can be suggested for this more general case.

- We have given some numerical examples where in the performance of the *globally optimal routing strategy* for the larger networks are compared with that of the suboptimal strategy which is synthesised from *locally optimal strategies* for the constituent units. Here again, an investigation on how good such a suboptimal strategy is, in comparison with the optimal one, in terms of bounds would make the study complete.

To summarize, this thesis has addressed to the issue of synthesising optimal and/or suboptimal routing strategies in some simple network units and in network topologies which are composed of these units. Although this work can be viewed more as theoretical exercise, we do consider that the investigations carried out in this thesis and the results obtained, are not without practical implications.

References

[1] Mischa Schwartz, Thomas. E. Stern, "Routing Techniques Used in Computer Communication Networks", *IEEE Transactions on Communications* , vol. Com-28, No: 4, April 1980.

[2] K.R.Krishnan and T.J.Ott, "Forward-Looking: A New State-Dependent Routing Scheme", *Proceedings of 12th ITC*, Torino, 1988.

[3] P.Chemouil, M.Lebourges, P.Gauthier, "Performance Evaluation of Adaptive Traffic Routing in a Metropolitan Network: A Case Study", *GLOBECOM' 89*, Dallas, 1989.

[4] E.Granel, J.L.Martinez, J.Lopez, S.Garcia, S.Herranz, I.Lapayese, "A Comparative Study of Several Dynamic Routing Algorithms with Adaptive Preselection and Selection Phases", *Proceedings of ITC-13*, 1991.

[5] A.Ephremides, P.Variaya, J.Walrand, "A Simple Dynamic Routing Problem", *IEEE Transactions on Automatic Control*, vol. AC-25, No.4, August 1980.

[6] L.Kleinrock, *Queueing systems Vol I: Theory*, Wiley Interscience, 1975.

[7] L.Kleinrock, *Queueing systems Vol II: Computer Applications*. Wiley Interscience, 1976.

[8] K.S.Narendra and M.A.L.Thatachar, "On the Behaviour of a Learning Automaton in a Changing Environment with Application to Telephone Traffic Routing", *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-10, No.5, May 1980.

[9] T.Yum and M.Schwartz, "The Join-Biased-Queue Rule and its Application to Routing in Computer Communications Networks", *IEEE Transactions on Communications*, vol. COM-29, pp 505-511, April 1981.

[10] G.Foschini and J.Salz, "A Basic Dynamic Routing Problem and Diffusion", *IEEE Transactions on Communications*, vol. COM-26, No.3, pp. 320-327, March 1978.

[11] F.Moss and A.Segall, "An Optimal Control Approach to Dynamic Routing in Networks", *IEEE Transactions on Automatic Control*, vol. AC-27, pp 329-339, April 1982.

[12] P.E.Sarachik and U.Ozguner, "On Decentralised Dynamic Routing for Congested Traffic Networks", *IEEE Transactions on Automatic Control*, vol. AC-27, pp 1233-1238, December 1982.

[13] C.Agnew, "Dynamic Modelling and Control of Congestion-Prone Systems", *Operations Research*, 24, 3(1976), 400-419.

[14] L.Kleinrock, *Communication Nets: Stochastic message flow and Delay*. New York: Mcgraw Hill, 1964.

[15] B.R.Hurley, C.J.R.Seidl, and W.F.Sewell, "A Survey of Dynamic Routing Methods in Circuit-Switched Traffic", *IEEE Communications Magazine*, Sept 1987 vol 25. No:9.

[16] K.R.Krishnan, "Adaptive State-Dependent Traffic Routing Using On-line Trunk-Group Measurements", *Teletraffic and Data Traffic in a period of change*, ITC-13, 1991.

[17] G.R.Ash, "Use of a Trunk Status Map for Real-Time DNHR", *Proceedings of 11th International Teletraffic Congress*, Kyoto, September 1985.

[18] R.J.Gibbens, F.P.Kelly and P.B.Key, "Dynamic Alternate Routing Trial", *Proceedings of the 12th International Teletraffic Congress*, Torino, June 1988.

[19] K.Mase and H.Uose, "Consideration on Advanced Routing Schemes for Telecommunication Networks", *Proceedings of 12th International Teletraffic Congress*, Torino, June 1988.

[20] L.Fretta, M.Gerla and L.Kleinrock, "The Flow Deviation Method: An Approach to Store-and-Forward Communication Network Design", *Networks*, Vol.3, pp 97-133,1973.

[21] Janusz Filipiak, *Modelling and Control of Dynamic Flows in Communication Networks*, Springer-Verlag, 1988.

[22] L.S.Pontryagin, V.G.Boltyanski, R.V.Gamkrelidze and E.F.Mischenko, *L.S. Pontryagin's selected works. Vol 4 : The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York, 1962.

[23] Michael Athans and Peter L. Falb, *Optimal Control: An Introduction to the Theory and its Applications*, McGraw-Hill, New York, 1966.

[24] *The NAG Fortran Library Manual, Mark 14*, The Numerical Algorithms Group Limited, 1990.

[25] Jeremiah. F. Hayes, *Modelling and Analysis of Computer Communications Networks*, Plenum Publishing Corporation, New York, 1984.

[26] E.Brockmeyer, H.L.Halstrom, and A.Jensen, "The Life and Works of A. K. Erlang", *Transactions of the Danish Academy of Technical Science ATS*, No.2 (1948).

[27] A.K.Erlang, "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges", *PO Electrical Engineering Journal*, 10, pp 189-197, (1917-1918).

[28] Feller.W, *Probability Theory and its Applications*, I-II, Wiley, 1966.

[29] Andre' Girard, *Routing and Dimensioning in Circuit-Switched Networks*, Reading MA: Addison-Wesley Publishing Company, 1990.

[30] Rene L.Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation", *IEEE Transactions on Information Theory*, vol. 37, No.1, pp 114-131, January 1991.

[31] Rene L.Cruz, "A Calculus for Network Delay, Part II: Network Analysis", *IEEE Transactions on Information Theory*, vol. 37, No.1, pp 132-141, January 1991.

[32] Clarke GM, "Use of Polya Distributions in Approximate Solutions to Nonstationary M/M/s Queues", *Communications ACM 24:206-217*, 1981.

[33] Moore SC, "Approximating the Behaviour of Nonstationary Single Server Queues", *Operations Research*, 1011-1032, 1975.

[34] Kendall, D.G, "Some Problems in the Theory of Queues", *Journal of Royal Stat. Soc*, B13 (1951), 151-185.

[35] Kobayashi H, "Application of the Diffusion Approximation to Queueing Networks: I. Equilibrium Queue Distributions", *JACM 21* : 316-328,1974.

[36] Kobayashi H, "Application of the Diffusion Approximation to Queueing Networks: II Nonequilibrium distributions and applications to computer modelling", *JACM 21:* 459-469, 1974.

[37] C.Rosenberg, F.Guillemin, R.R. Mazumdar, "On Quantile Measures for Traffic Characterisation", *Proceedings of ITC Sponsored Seminar on Teletraffic Analysis Methods for Current and Future Telecom Networks*, Bangalore, Nov. 1993.

[38] K.S.Narendra, A.Wright and L.G.Mason, "Application of Learning Automata to Telephone Traffic Routing and Control", *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-7, No.11, November 1977.

[39] P.R. Srikantakumar, K.S.Narendra, " A Learning Model for Routing in Telephone Networks", *SIAM Journal on Control and Optimization*, vol 20, No.1, January 1982.

[40] Mischa Schwartz, *Computer Communication Network Design and Analysis*, Prentice-Hall, Inc., Englewood Cliffs, New Jersy 07632.

[41] H.Maisel, G.Gnugnoli, *Simulation of Discrete Stochastic Systems*, Science Research Associates Inc, Chicago, 1972.

[42] R.T. Proser, "Routing Procedures in Communication Networks", *IRE Transactions on Communication Systems*, CS-10, No.4, 1962. "Part I: Random Procedures", 322-329. "Part II: Directory Procedures", 329-335.

[43] P.Chemouil, J.Filipiak, P. Gauthier, " Performance Issues in the Design of Dynamically Controlled Circuit-Switched Networks", *IEEE Communications Magazine*, October 1990.

[44] K.R.Krishnan, "Network Control with State-Dependent Routing", *Proceedings of the ITC Special Seminar.*, Adelaide, Australia, 1989.

[45] P.Gauthier and P.Chemouil, " A System for Testing Adaptive Traffic Routing in France", *Proceedings of GLOBECOM '87*, Tokyo, Japan, 1987.

[46] G.R.Ash, R.H.Cardwell and R.P.Murray, "Design and Optimization of Networks with Dynamic Routing", *Bell Systems Technical Journal*, vol.60, no.8, 1981.

[47] G.R.Ash *et al.*, "Robust Design and Planning of a Worldwide Intelligent Network." *IEEE Journal on Selected Areas in Communications.*, vol.7, no.8, 1989.

[48] H.Akimaru and K.Kawashima, *Teletraffic Theory and Applications*, Springer-Verlag, 1993.

[49] Dimitri Bertsekas and Robert Gallager, *Data Networks*, Second Edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., U.S.A.

[50] T.B. Crabill, D. Gross, M.J. Magazine, "A Classified Bibliography of Research on Optimal Design and Control of Queues", *Operations Research*, vol.25, No. 2, March-April 1977.